Choice experiments:
An approach to assess recreational values in an ecological thresholds framework

By Catalina M. Torres Figuerola

March 2007

PROJECT PROPOSAL

Directed by Antoni Riera Font

DEPARTAMENT D´ECONOMIA APLICADA

UNIVERSITAT DE LES ILLES BALEARS

INDEX

ACKNOWLEDGEMENTS

I thank Nancy Bockstael, Professor at the University of Maryland, and Robert Hicks, Professor at the College of William and Mary of Williamsburg, Virginia, for all what they have taught me. I have benefited greatly from their knowledge in economic valuation. Very especially, I thank Antoni Riera, my PhD director and fellow traveler in this process, for his help, his patience, and his great dedication and commitment. This project proposal would not exist without him.

I. INTRODUCTION

Over the last years, sustainable development, global change and ecosystems related topics have captured the attention of researchers. The high amount of research projects funded by the European Commission (EC) under the Sixth Framework Programme *Sustainable Development, Global Change and Ecosystems* mirrors the nature of priority assigned to these research areas. No doubt, there is a need for sustainable policies. Strategies implemented under sustainability issues must be understood as the integration of environmental, economic and social aspects. Despite growing recognition of the importance of ecosystem functions and services, they are often taken for granted and overlooked in environmental decision making. But the increasing anthropogenic pressure over the environments can lead ecosystems to exceed their carrying capacity and experience regime shifts between alternative stable states that could become irreversible. For this reason, sustainable strategies must rely on the identification of target values associated to environmental pressures and impacts, that is, of thresholds that must not be exceeded if sustainable development wants to be achieved.

In this context, there must be recognition of the potential for conflict in decision making processes involving choices between the conservation and restoration of ecosystems and the expansion of human activities. In this sense, the role of environmental economists is essential. It is necessary to know the economic value of the ecosystems goods and services so that they can be compared with the economic value of activities that may compromise them and so that improvements to one ecosystem can be compared to those in another. However, the challenge for environmental economists is to assess these values in a framework of ecological thresholds and possible irreversibilities. In other words, a good valuation of ecosystem goods and services implies the integration of ecology and economics if sustainability must be achieved. Ecological discontinuities affect the goods and services provided by ecosystems, which is supposed to have some influence on the utility function of individuals. In this sense, the success of the integration of ecology and economics depends on the good interpretation of research on ecosystem functions by ecologists so that *service-level information* can be communicated to economists.

The analysis of ecological thresholds to serve as input for environmental policies is at the core of the EC-funded *Thresholds of Environmental Sustainability* project. It is focused on developing, improving and integrating research tools and methods to guide the implementation of sustainable strategies to be applied in the European coastal zone and, hence, to face the challenges contained in the EU Strategy on Sustainable Development, the Johannesburg Summit on Sustainable Development and the UN Millennium Development Goals. Within the framework of this project, this work wants to stress the environmental degradation of coastal waters and its effects on water related recreational values. Human activities, high population growth, and industrial, commercial, tourism, and residential development have led to more pollution and, hence, perturbation of the dynamics of marine ecosystems, which has negatively affected the health of coastal waters and even changed, in many cases, their color and transparency appearance. One of the effects of the growing human pressure over the coast is the increase in nutrients emissions into the waters, especially through treated waste water flows and related biodegradable effluents, and agricultural fertilizers. All this contributes to accelerate coastal water quality degradation through what is called *eutrophication process*. Following the words of the Urban Waste Water Treatment Directive (91/271/EEC)[1], 'eutrophication means the enrichment of water by nutrients, especially compounds of nitrogen and/or phosphorus, causing an accelerated growth of algae (i.e. *algal blooms*) and higher forms of plant life to produce an undesirable disturbance to the balance of organisms present in the water and to the quality of the water concerned'. Given specific environmental conditions, as a low water renewal, coastal waters can experience a change in their ecological status when a certain level (i.e. *threshold level*) of nutrients concentration is achieved. Ecological disturbances may lead to an abrupt, and maybe substantial, disruption (i.e. *threshold effect*) in the supply of one or more coastal water services. In this context, it is expected a change in the utility function of coastal water users, in the sense that the value assigned by them to a water service in a context of thresholds effects must not be the same as the one assigned in a context of a good ecological status of coastal waters.

Two of the consequences of eutrophication are the loss of water transparency and the change in water color due to the proliferation of algal blooms. The high inversely correlation between these two features of coastal water degradation and environmental aesthetics makes water recreational values interesting values to be assessed in a context of ecological discontinuities, because aesthetics is supposed to highly influence them. Furthermore, water related recreation represents a large component of the total benefits from water quality improvements and its evaluation becomes an important part of a *cost-benefit analysis* (CBA) to assess policies affecting marine water quality in general (Freeman, 1982). In this context, the marginal recreational benefits associated with a particular service provided by coastal waters

[1] Council Directive 91/271/EEC of 21 May 1991 concerning Urban Waste Water Treatment.

may either be fairly constant or change in a fairly reasonable predictable manner with the provision of that service, but once the threshold is reached, with subsequent thresholds effects in terms of loss of water transparency and/or change in water color, not only may there be a large *jump* in the value of that service, but even how the supply of the service changes may be less predictable.

However, complexity involved in ecosystems dynamics and uncertainty surrounding both the magnitude and the timing of any threshold effect associated to an ecological disturbance makes the valuation of ecosystems services a difficult task. Among the economic valuation techniques, stated preference (SP) methods are viewed as the most suitable ones for measuring values in a framework of ecological non-linearities. This is because SP methods, in contrast with revealed preference (RP) techniques, do not infer people's preferences through the observation of their behavior in real market situations, that is, do not need people to have made choices in response to thresholds effects in the past, which is in line with the unpredictable thresholds effects occurrence. Thus, it seems to be reasonable that the valuation technique of interest be designed to value a variety of plausible ecosystem scenarios so that it can be described the sensitivity of the obtained values to each possible outcome. One of the SP methods that could play an important role in this topic is the *choice experiment* (CE) due to its design requirements involving the construction of different choice sets with a specific number of different scenarios (i.e. alternatives) to be presented to respondents in order for them to choose their most preferred one from each choice set.

In this context, the purpose of this work consists of doing an in-depth analysis of the basic issues underlying the CE method to make progress with the examination of its potential to assess economic values in a context of ecological discontinuities. For it to be accomplished, the technique will be first presented from its origins to its applicability in environmental economics. This will allow having a good idea of the role of CEs in valuation research at the time that will show some of its advantages and drawbacks over other methods. The knowledge acquired in this first task and the further study of the main CE methodological issues will contribute to a better understanding of the environmental economics literature concerning CEs, whose revision will allow the identification of the existing gaps.

The structure of this work is as follows. In the next section, an analysis of the evolution of CEs from its emergence in marketing research to its adoption by environmental economists is done. In section III, the basic methodological issues regarding model specification are described. Section IV is based on an examination of the main ideas underlying the experimental design process. A review of the environmental literature involving CEs is carried out in section V. Then, a research line focused on the capability of the CE to deal with economic valuation in a thresholds framework is presented in section VI. Bibliographic references used are listed in the section VII. Finally, a table summarizes the most common issues characterizing the CE applications carried out by environmental economists over the last decade.

II. CHOICE EXPERIMENTS: FROM MARKETING TO ENVIRONMENTAL ECONOMICS

In an economy, goods and services become important for individuals as long as they contribute to their human wellbeing. In other words, people assign an economic value to goods and services that matter to them irrespective of whether they have a monetary value or not. In this context, this value must be taken into account in economic decision making if socially efficient solutions want to be achieved. However, evidence shows that resource allocation decisions are mostly made with information on the monetary value of marketed goods and services. Those that traditionally have not been assigned a price tend to be overlooked. The inexistence of a price for non-market goods and services, and the consequent difficulty to deal with them in decision-making, is viewed as one of the reasons that explain the lack of attention they have received. Nevertheless, the high relevance that many of them have gained over the last decades, especially those provided by natural resources, justifies the need to solve this market failure. Economic valuation techniques have emerged as a way to assign a monetary value to non-market goods and services to allow their integration into economic analysis.

SP methods constitute one of the two broad categories of non-market valuation techniques. They rely on asking people to state their preferences for alternative hypothetical scenarios through the use of survey techniques. Their most basic form, and one of the most commonly used approach, is the *contingent valuation method* (CVM), whose origins can be found in Ciriacy-Wantrup (1947), although the first study is done by Davis (1963). CVM involves administering a survey to respondents in which a hypothetical quality change is described. The elicitation of people's willingness to pay (WTP) for the change by asking respondents hypothetical questions can be done through the use of different response format approaches (i.e. elicitation formats) that can lead directly to WTP or provide information to estimate the preferences. The fact that researchers can dictate hypothetical scenarios allows SP methods to be adapted to handle just about any valuation problem. This helps to explain the wide use of CVM in valuation research (Carson, 1998).

In parallel with the use of the CVM, another kind of SP methods is developed. They include *rating* applications, *ranking* exercises, CEs and *paired comparisons*. All of them share the same theoretical foundation based on Lancaster (1966)'s work on the study of product demand through a microeconomic analysis of products characteristics. Lancaster shows that the utility of individuals is derived from the bundle of attributes that, in fixed proportions, define a good or a situation, in contrast with the traditional approach that goods are the direct objects of utility. After the emergence of Lancaster's work, a new technique is implemented within the framework of the mathematical and experimental psychology. It is called *conjoint measurement* and consists of decomposing overall judgments regarding a set of alternatives into the sum of weights on attributes of these alternatives.[2] This method is rapidly adopted by marketing researchers in the 1980's. Since then, it has been commonly known as *conjoint analysis* (CA). In words of Batsell & Louviere (1991), 'conjoint analysis refers to a family of paradigms for the algebraic representation of individual judgments of multiattribute stimuli'.

Traditionally, the concept of CA has been related to rating exercises but, in reality, CA techniques are viewed as a broader approach including not only rating applications but also ranking studies, CEs and paired comparisons (Bateman et al., 2002; Hanley et al., 2001). Survey respondents are presented with a number of alternatives, described by different attributes at different levels. In a rating exercise, individuals are asked to rate the alternatives individually in a semantic or numeric scale. In a ranking study, respondents are required to rank the set of the available alternatives. In a CE, alternatives are grouped into different choice sets and individuals are asked to choose their most preferred one from each choice set. When the application is based on paired comparisons, respondents have to choose their most preferred alternative out of a set of only two at the same time that they are also required to indicate the strength of their preference in a semantic or numeric scale. All of these CA techniques share common features. First, they require the identification of key attributes that underlie the preferences of respondents for different alternatives. Second, they use experimental design or other methods to obtain scenarios. Experimental design implies the use of statistical design theory, whose main goal is to obtain orthogonal designs[3] that allow the construction of choice sets or scenarios in a way that parameter estimates are not confounded by other factors. Third, they utilize statistical methods to decompose the preferences into components due to each attribute level. Fourth, they allow prediction of preferences or choices through the use of simulation methods (Louviere, 2001).

However, despite these similarities, the rating approach presents two important differences with respect to the other CA techniques. On one hand, it does not involve a direct comparison of alternatives. On the other one, strong assumptions about the cardinality of rating scales must be done in order to transform ratings into utilities. In

[2] The term *conjoint* means that a bundle of attributes are considered jointly.

[3] Orthogonality is a mathematical constraint requiring that all attributes be statistically uncorrelated, that is, independent of one another.

contrast with the other CA techniques, this implies a departure of rating experiments from contexts of choice actually faced by consumers. More specifically, it does not rely on the economic theory framework based on random utility maximization (RUM) models and, hence, it is not linked to economic choices. This makes it a doubtful method for consistent welfare estimations (Hanley et al., 2001; Morrison et al., 1996). RUM models have their basis in the behavioral theory evolved out of Thurstone (1927)'s work. From a psychological point of view, Thurstone develops a law of comparative judgment in an attempt to explain choices of individuals in paired comparison situations. His ideas suppose the first step to model choice decisions (Hanley et al., 1998b; Louviere, 2001)and become the conceptual foundation of the discrete choice theory as formulated by McFadden (1974) for economic analysis. This theory is based on the idea that individuals choose the alternative that maximizes their utility from a set of available alternatives. Starting with Luce (1959)'s choice axiom about the independence of irrelevant alternatives (IIA), as linked by Marschak (1960) to the random utility model of Thurstone, McFadden develops an econometric model, the conditional logit (CL) model, or, less correctly, the multinomial logit (MNL) model, that combines hedonic analysis of alternatives, described by their characteristics or attributes, and the random utility maximization.

The lack of a RUM basis in rating experiments explains the emergence of other techniques within the CA approach, which have been adopted, since their origins, not only by marketing researchers but also by transportation practitioners. One of them is the ranking method. However, it also presents some disadvantages that questions its ability to elicit preferences. Literature has identified some of these problems. The difficulty involved in making interpersonal comparisons of ranking data is one of them, also shared by ratings experiments. Furthermore, ranking alternatives becomes a difficult task for respondents, especially when the number is large, with many attributes and levels. On the other side, rankings exercises can be viewed as a series of choices in which respondents face a sequential process whereby they must first choose their most preferred alternative from the available set of alternatives. Once they have made a choice, they must choose their most preferred alternative from the remaining set. And so on. This choice process makes possible that a *status quo* (SQ) alternative be not present in the choice set. The fact that respondents can be forced to choose between alternatives from a set where there is no a baseline scenario implies an inconsistency with welfare economics assumptions of utility maximization and demand theory. For all these reasons, at the same time that rating and ranking applications have been carried out by some researchers, other authors have moved their attention towards the CE approach.

It seems to be that the term *choice experiment* is first used in transportation economics by Louviere & Hensher (1982). The authors propose the CE as the methodology adequate to satisfy statistical conditions for a variety of econometric choice models and useful for making forecasts to test external validity. One year later, Louviere & Woodworth (1983) present the CE approach in the marketing field as a method capable of studying choice under controlled conditions, which allows studying aggregate consumer choice behavior and attributing trade-offs process in choice. In a CE, survey respondents are presented with a set of alternatives grouped into different choice sets, described by different attributes at different levels, and asked to choose their most preferred one from each choice set. The sequence of choice results enables the probability of an alternative being chosen to be modeled in terms of their attributes. According to that, it is expected that the higher the level of a desirable attribute in an alternative, other factors being equal, the greater the utility associated with that option and the more likely its choice by the respondent. And viceversa, the more of an undesirable attribute in an alternative, the lower the utility and the less likely its choice. Such models allow researchers to find out which trade offs respondents make between the attributes and their responses to different scenarios. In Bennett & Adamowicz (2001)'s words, 'by observing and modeling how people change their preferred option in response to the changes in the levels of the attributes, it is possible to determine how they trade-off between the attributes'. According to these features, marketing and transportation researchers have viewed the CE as an excellent technique to predict market shares in situations when a new product is launched or a change in an existing one happens, and when alternative modes of transport are considered, respectively.

The first CE application to environmental management problems is carried out by Adamowicz et al. (1994).[4] Originally, the interest shown by environmental economists in the use of CEs can be explained by the understanding of the method as a good way to obtain the necessary data for resource allocation in situations where no market exists. The possibility of eliciting the WTP of respondents to move from the SQ scenario to another one representing the result of a policy about which people's preferences want to be estimated allows the outcomes of CEs to be used as inputs of a CBA of alternative policies. But it is the numerous advantages the technique presents over other SP methods, especially over the widely used CVM, what has motivated the

[4] Within the CA approach, environmental economists have first focused their attention on ranking experiments. The first ranking study is Rae (1983)'s work to value visibility impairments at Mesa Verde and Great Smoky Mountains National Parks. Rating exercises and CEs have started to be applied simultaneously in environmental valuation. Mackanzie (1993) is who first carries out a rating application by making a comparison between contingent preference methods. Even health economists have increased their interest in applying CEs, whose potential has been emphasized by some authors (Hanley et al., 2003).

growing use of CEs in environmental valuation research since the 1990's. In this sense, the ability of the method for estimating a financial indicator of the WTP for one additional unit of a non-monetary attribute (i.e. *implicit price* or *attribute value*) has been argued to be one of the most important advantages of the CE over the CVM. Moreover, CEs are not only appropriate methods to elicit *passive use* values[5], as they allow asking respondents about their choices of environmental quality settings, but also to estimate *use values*, because CEs can be used to modify the levels of the environmental quality, allowing their expansion beyond the current ones, and to identify the value for each specific change. Another advantage of the CE over the CVM is given by the possibility to obtain each possible outcome in case of uncertainty of attribute levels, whereas the CVM only permits to obtain one value for an expected quality change (Garrod & Willis, 1999; Hanley et al., 2001; Hanley et al., 1998b).

One of the major critics to SP methods, especially to the CVM, is they are very prone to give biased results due to their hypothetical nature (Diamond & Hausman, 1994; Mitchell & Carson, 1989). In this sense, CEs are argued to minimize some of the most common potential biases related to CVM, as *protest bids*, *strategic behavior*, *yeah-saying* (i.e. *warm glow effect* or *compliance bias*) and *ethical protesting*. This minimization is achieved by forcing respondents to choose one alternative from the set of the available ones for a sequence of choice sets. Furthermore, the sequential process allows the elicitation of more information to be used in the data analysis (Birol et al., 2006a; Holmes & Adamowicz, 2003). CEs can also reduce the *embedding problem* (i.e. *insensitivity to scope*) encountered in the CVM. It is because CEs allow for internal consistency tests, in the sense that models can be fitted on subsets of the data. In other words, scope tests are built into CEs (Alpízar et al., 2001; Hanley et al., 1998b).[6] Sometimes CVM studies pass the scope test. According to Bateman (2002), 'CVM studies tend to find that values are higher for higher quantities'. However, in these studies it is difficult to find scope tests that allow observing WTP values across a wide range of quantities. Observing that is easier in CE exercises, in which scope can be an attribute itself and more combinations of WTP and quantity can be addressed. Another advantage of CEs over the CVM has to do with the cost of the valuation study. Some authors argue that 'CEs can be less expensive due to their ability to value program attributes in one single questionnaire and because they are more informative than discrete choice CVM surveys' (Hanley et al., 2001). In addition, it is said that CEs are

[5] Adamowicz et al. (1998a) carry out the first CE application estimating passive use values.

[6] Hanley et al. (2001) point out that one of the methods adequate to assess sensitivity to scope consists of making a *within group* or an *internal test*, that is, presenting each individual with a number of valuation scenarios that differ according to scope. Thus, scope tests are built into CEs because this way of proceeding constitutes one of its basic issues.

better than the CVM in terms of benefit transfer (BT) as well as in terms of modeling substitution possibilities (Boxall et al., 1996; Hanley et al., 1998b; Rolfe et al., 2002).

CEs are also argued to have advantages over RP methods. The possibility to introduce or remove either attributes or attribute levels allows more control over the experimental design, in contrast with the case of real market situations (Adamowicz & Boxall, 2001; Holmes & Adamowicz, 2003). On one hand, this permits the introduction of attributes associated with passive use values and, on the other one, it enables the inclusion of wider attribute levels than the ones found in real contexts (Carson et al., 1994). Furthermore, manipulation of attributes and their levels is useful, because many policy decisions are concerned with changing attributes levels and not with gaining or loosing the environmental good as a whole. Another advantage of CEs over RP methods is given by the fact that RP data usually present, as well as lack of variation, collinearity between the explanatory variables, which makes difficult the estimation of marginal values of attributes. On the contrary, the control over the design matrix in CEs can eliminate this collinearity, unless explicitly built into the design, allowing a greater statistical efficiency. Within the optimistic framework of CE capabilities, some researchers have attempted to answer the question about whether choices observed in CEs reveal the same information about preferences than the ones observed in parallel from RP data sources. As a result of a literature review, it has been found a positive answer (Louviere, 2001).

However, the method also presents some drawbacks. Hanley et al. (1998b) state two types of problems related to the description of an environmental good in terms of its attributes. The first one is referred to the fact that individuals, due to the complexity of their perceptions, can view the asset different from the described one, and, hence, consider either more, less or other attributes than the ones used by the researcher. This also leads to the point that *not always the whole is equal to the sum of the parts* (Bateman et al., 2002). The second one is related to the violation of ecosystems dynamics, in the sense that orthogonality between attributes does not allow the use of attributes where the existence of one of them depends on the previous existence of the other one, which is a very frequent phenomenon in ecosystems functioning. In other words, the existence of *causally related attributes* is not possible. However, this is really not a bad thing, because if they were included in the design, respondents might spend a lot of time trying to understand the causal relations to assign a greater meaning to the alternatives and potentially simplify the decision making process, which would likely lead to biased results (Blamey et al., 2001). On the other side, Hanley et al. (2001) point out the sensitivity of SP methods to study design and the fact there is no reason to believe that CEs solve hypothetical biases. They also suggest that the valuation of the sequential provision of goods in multiattribute programs is probably better undertaken by the CVM, because values for a sequence of elements implemented by a policy can be easily derived from the application of this method. As an additional disadvantage, it must be taken into account that repeated answers per individual could cause statistical problems in terms of possible correlation between them (Adamowicz et al., 1998b).

From a psychological and experimental economics perspective, CEs also present some weaknesses. It has been argued the existence of a *cognitive difficulty* related to complex choices between alternatives, in the sense that when choices are complex, respondents use heuristics and rules of thumb to simplify decision tasks. It has been stated the existence of *learning and fatigue effects* when carrying out a CE, effects that can lead to irrational choices. In this context, Mazzota & Opaluch (1995) study the effect of increasing the number of attributes that vary between alternatives and found that where four or more attributes are varying, respondents consistently eliminate one or more attributes from their consideration in order to reduce task complexity. As a conclusion, they state that increased complexity leads to increased random errors. Swait & Adamowicz (1996) observe an effect of task complexity on taste parameter estimates by finding out an inverted U-shape relationship between choice complexity and variance of underlying utility amounts. Bateman et al. (2002), on the other hand, have found significant insensitivity to scope in separate CEs when respondents are given too many choice sets.

Despite all this, proponents of CEs are optimistic regarding solving the drawbacks cited above. In fact, new CE design issues are being developed and tested in an attempt to minimize its weaknesses. On the other side, the high array of advantages offered by the CE, especially that of design control, along with the ability of SP methods to handle any valuation problem due to their hypothetical nature, makes the method very attractive to be applied in the assessment of economic values. In *an examination of the historical trends and future directions in environmental valuation*, Adamowicz (2004) states that 'the most important advance in the area of preference elicitation is a movement towards the analysis of individual level data using RUM models, accompanied by an increasing interest in both behavioral and experimental economics and understanding individual choice behavior'. When this research trend is combined with the need of integration of ecology and economics to take into account non-linearities in ecosystem dynamics, then a double challenge emerges for environmental economists. First, overcoming the CE weaknesses and enhancing its strengths. And second, using CEs to assess successfully economic values in a context of ecological thresholds effects.

III. MODEL SPECIFICATION

Understanding contexts of choice faced by consumers requires considering issues underlying RUM models. The individual choice behavior problem based on maximizing the utility converts the choice problem into a distribution of behavioral responses that, along with some specific axioms, makes the problem more tractable by analysts at the same time that allows obtaining consistent welfare estimations. RUM models are used to specify models of behavior both in RP studies and in SP applications. Their integration into the category of SP methods is oriented toward linking hypothetical market situations to real economic choices in an attempt of enhancing the potential of SP techniques to elicit preferences. A good example of this is the CE.

In choice problems specified by RUM models, an individual *i* is presented with a set of *J* mutually exclusive alternatives and asked to choose their most preferred one. It is supposed that he maximizes his utility and, hence, after evaluating each and every alternative in the choice sets, chooses the alternative *j* that gives him the highest utility level. In this context, it is assumed that the overall utility U_{ji} associated with an alternative *j* for an individual *i* is given by the sum of two parts: a *deterministic* or *systematic component* V_{ji} (i.e. *representative utility*) and a *random* or *stochastic component* ε_{ji}, as is shown below:[7]

$$U_{ji} = V_{ji} + \varepsilon_{ji}\,; \quad \forall j \in \; j = 1,\ldots,j,\ldots,J; \quad \forall i \in \; i = 1,\ldots,i,\ldots,N \qquad (1)$$

The disgregation of U_{ji} into the two components is due to recognition by the analyst of the existence of other individual-specific utility influences different from the ones identified by him.[8] In this sense, V_{ji} is the part of utility associated with the observed factors influencing it, whereas ε_{ji} represents the unobserved sources of utility. These unobservables can be characteristics of the individuals and/or attributes of the item, and they can stand for both variation in preferences among members of a population and measurement error (Hanemann & Kanninen, 1999). In this context, given a set of *J* mutually exclusive alternatives, it is assumed that an individual *i* selects an alternative *j if and only if* the utility U_{ji} that it gives to him is greater than (or equal to) the one associated with an alternative $k \neq j$ in the same choice set. That is:

[7] This function is called *conditional indirect utility function*, because the utility is conditional on the choice of alternative *j*.

[8] It is to recall that individuals know their preferences with certainty and do not consider them stochastic.

iff $U_{ji} \geqslant U_{ki}\ ; \quad k \neq j$ (2)

iff $V_{ji} + \varepsilon_{ji} \geqslant V_{ki} + \varepsilon_{ki}\ ; \quad k \neq j$ (3)

The goal of a RUM choice model is to identify the attributes that affect the utility of individuals and estimate their significance. For this reason, it is necessary to specify a functional form for V_{ji}. It is usually used a linear in parameters, additive form that, following the words of Louviere et al. (2000), 'represents the composition rule that maps the multidimensional attribute vector into a unidimensional overall utility'. It is written as:

$$V_{ji} = \beta_{0ji} + \beta_{1ji} f\left(X_{1ji}\right) + \ldots + \beta_{qji} f\left(X_{qji}\right) + \ldots + \beta_{Qji} f\left(X_{Qji}\right); \quad q = 1, \ldots, q, \cdots, Q \qquad (4)$$

where X_{qji}'s represent the observed attributes that affect utility of alternative *j* and individual *i* and β_{qji}'s are their parameter weights. Parameter β_{0ji} is called *alternative-specific constant* (ASC). It is not associated with any of the observed and measured attributes and represents, on average, the role of all the unobserved sources of utility. In other words, it captures the mean effect of the unobserved factors in the error terms for each alternative. The notation $f\left(X_{qji}\right)$ means that attributes can enter the utility function in different ways. However, to facilitate the analysis, V_{ji} will be considered not only additive, linear in parameters, but also linear in attributes, as shows the next expression:[9]

$$V_{ji} = \beta_{0ji} + \sum_{q=1}^{Q} \beta_{qji} X_{qji} \qquad (5)$$

The uncertainty derived from the random component of the utility expression leads to deal with the choice problem in terms of probabilities, that is, only statements in terms of probability can be made. Then, the individual behavioral choice rule available to the analyst implies that the probability P_{ji} of individual *i* choosing alternative *j* from a set of *J* mutually exclusive alternatives is equal to the probability that the utility of alternative *j* be greater than (or equal to) the utility associated with alternative *k* after evaluation of each and every alternative in the choice set. Thus, the choice problem becomes as follows:

[9] There exists the possibility to specify complex non-linear forms. However, for simplicity reasons, only the linear form is analyzed in this work. In this context, when a variable does not vary across alternatives for individual *i*, as the socio-demographic characteristics (SDCs), it can not be included separately into the specification of V_{ji}. On the contrary, it must enter the expression by interacting with some other alternative-specific attribute.

$$P_{ji} = P\left(U_{ji} \geqslant U_{ki} \mid k \in J \right) \quad \forall k \neq j \tag{6}$$

$$P_{ji} = P\left(V_{ji} + \varepsilon_{ji} \geqslant V_{ki} + \varepsilon_{ki} \mid k \in J \right) \quad \forall k \neq j \tag{7}$$

$$P_{ji} = P\left(V_{ji} - V_{ki} \geqslant \varepsilon_{ki} - \varepsilon_{ji} \mid k \in J \right) \quad \forall k \neq j \tag{8}$$

For the individual choice model to be more tractable, it has been developed some axioms aimed at facilitating the interpretation of the empirical results of selection probabilities. The main used one is the *IIA axiom* postulated by Luce (1959). It states that the ratio of the probabilities of choosing one alternative over another (given that both alternatives have non-zero probability of choice) is unaffected by the presence or absence of any additional alternative in the choice set. The IIA property implies that the error terms within the utility specification of the alternatives are identically distributed and independent across them. This *independently and identically distributed assumption* is commonly known as the *IID assumption* and is equivalent to the *constant variance assumption*, because an identical distribution implies a same variance. On the other hand, the independence across alternatives indicates *covariances* (i.e. cross-related terms) set to zero. In this sense, the *variance-covariance matrix* that describes the full correlation structure between J alternatives is given by:

$$\begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1j} & \cdots & \sigma_{1J} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{j1} & \cdots & \sigma_{jj}^2 & \cdots & \sigma_{jJ} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{J1} & \cdots & \sigma_{Jj} & \cdots & \sigma_{JJ}^2 \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 & 0 \\ 0 & \sigma^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2 \end{bmatrix} \tag{9}$$

However, it is common practice to normalize one of the variances to equal to *1*. Thus, the variance-covariance matrix for this simple case becomes as follows:[10]

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{10}$$

[10] The simplicity of the IID assumption allows analyzing the underlying properties of this kind of choice models easily. Nevertheless, this assumption can be relaxed to construct more complex models, especially in cases where there is concern about possible violation of constant variance and/or correlated alternatives. These models include the *multinomial probit*, the *nested logit*, the *random parameters logit model*, and the *heterogeneous extreme value logit.*

Among the high variety of statistical distributions available for these IID error terms, one commonly used is that of *extreme value (EV) type I*, also known as *Gumbel* distribution, which arises to the normal distribution. The essential difference between them is in the tails of the distribution, where the extreme value resides. In particular, this implies that the higher the number of alternatives in the choice problem, the more noticeable the difference between the two distributions. The expression for the *cumulative distribution function* (CDF) of the EV type I distribution is given by:

$$\text{CDF: } F(\varepsilon) = P(\varepsilon_j \leq \varepsilon) = \exp(-\exp{-\varepsilon}) = e^{-e^{-\varepsilon}} \qquad (11)$$

from which the *probability density function* (PDF) can be derived:

$$\text{PDF: } f(\varepsilon) = F'(\varepsilon) = \exp(-\varepsilon)\exp(-\exp{-\varepsilon}) = e^{-\varepsilon}e^{-e^{-\varepsilon}} \qquad (12)$$

Expression (11) treats the information as unobserved and randomly distributed across the unknown distribution represented according to (11). However, as earlier said, there is a part of the utility specification that represents observed information. Thus, factor ε can be replaced by other information, revealed in (8). In this sense, rearranging (8) to reflect condition (11) and making some transformations lead to the basic form of the CL model as developed by McFadden (1974), in which the probability of choosing alternative *j* by individual *i* from a set of *J* mutually exclusive alternatives can be written as:

$$P_{ji} = \frac{\exp(\mu V_{ji})}{\sum_{k=1}^{J}\exp(\mu V_{ki})} = \frac{1}{\sum_{k=1}^{J}\exp{-}\left(\mu(V_{ji} - V_{ki})\right)} \quad ; \quad k = 1, \ldots, j \ldots, J \qquad (13)$$

where μ is the scale parameter and is usually normalized to *1*, implying constant error variance.[11] Thus, expression (13) becomes:

[11] The true parameters are actually confounded with the scale parameter. In fact, the parameter estimates represent μβ's and not the true β's. In this context, it is impossible to identify the scale parameter from the data. The impact of the scale parameter on the estimated coefficients imposes restrictions on their interpretation. In this sense, all parameters within an estimated model have the same scale and, hence, their signs and relative sizes can be compared. However, it is not possible to directly compare parameters from different models as the scale parameter and the true parameters are confounded. Only, the comparison between estimated parameters from two different data sets or the combination of data sets is possible (Alpízar et al., 2001).

$$P_{ji} = \frac{\exp\left(V_{ji}\right)}{\sum_{k=1}^{J} \exp\left(V_{ki}\right)} = \frac{1}{\sum_{k=1}^{J} \exp\text{-}\left[\left(V_{ji} - V_{ki}\right)\right]} \quad ; \quad k = 1, \ldots, j \ldots, J \text{[12]} \tag{14}$$

There are diverse statistical techniques to estimate the parameters of the CL model. The most widely used is the *maximum likelihood (ML) estimation*, based on the idea that a given sample could be generated by different populations and is more likely to come from one population than another. The issues underlying this estimation approach lead to the construction of a probabilistic function *L* known as the *likelihood function* and represented by:

$$L = \prod_{i=1}^{N} \prod_{j=1}^{J} P_{ji}^{f_{ji}} \tag{15}$$

where f_{ji} is a dummy variable equal to *1* if alternative *j* is chosen by individual *i* and equal to *0* otherwise. In *L*, parameters β_{qji}'s within P_{ji} are variable and unknown, whereas variables X_{qji}'s are known and, hence, fixed. The maximization of (15) with respect to the parameters, by making $\partial L / \partial \beta_{qji} = 0$, allows obtaining the *maximum likelihood estimators* $\hat{\beta}_{qji}$'s (MLEs). These parameter estimates are the MLEs of the population parameters and represent the values that are most likely to have generated the sample of observed variables. They are invariant to monotonically increasing transformations of *L*. This makes possible to use the logarithm of *L*, that is, the *log likelihood function L**, to maximize the problem, which is easier from a mathematical point of view. Thus, the function to be maximized becomes:

$$L^{*} = \sum_{i=1}^{N} \sum_{j=1}^{J} f_{ji} \ln P_{ji} \tag{16}$$

MLEs are estimates of the weight of attribute *q* in the utility V_{ji} of alternative *j* for individual *i*. They are also known as *marginal utilities* or *part-worth utilities.*[13] By taking the values of X_{qji}'s for individual *i* and alternative *j* and the value of the parameter estimates, and substituting them in (5), an estimation $\hat{V}_{ji}$ of the representative utility is

[12] According to (14), it is easy to see that the variables that do not vary between alternatives, as the SDCs, do not affect the choice made by individuals.

[13] When they are negative, they represent values of *disutility*. On the other hand, when they are almost identical for some alternatives, they can be treated as *generic* parameters for these alternatives.

obtained. It is usually interpreted as the *relative utility* of alternative *j* to individual *i*, because what matters is the utility level associated with an alternative relative to that of another alternative in the same choice set. From (8), it can be seen that only differences in utility matter in choice models. This implies that the only parameters that can be estimated are those that capture differences across alternatives (Train, 2003). Therefore, there is some *base reference* against which utility of each alternative is compared.[14]

ML procedure allows calculating *asymptotic* standard errors for the $\hat{\beta}_{qji}$'s in the CL model and use them to test their statistical significance using *asymptotic t-tests*, that is, tests valid in only very large samples. To test the overall goodness-of-fit of the ML estimation procedure, two tests can be used. On one hand, the analyst can resort to the *likelihood ratio test*, which allows testing the contribution of particular subsets of variables to the utility specification, that is, whether the probability P_{ji} of individual *i* choosing alternative *j* is independent of the parameter values within the subset considered. On the other one, it can be used the *test of prediction success*, which involves a comparison of the summed probabilities from the models. In other words, it permits to compare the expected number of individuals choosing a specific alternative with the observed behavior for the sample.[15]

An advantage of using RUM models through CL is their usefulness to assess the effects of a lot of policies. In this context, it can be calculated both *direct* and *cross elasticities* of choice. The first ones indicate the percentage change in the probability of choosing alternative *j* with respect to a percentage change in an attribute of the same alternative, whereas the second ones mean the percentage change in the probability of choosing alternative *j* given a percentage change of an attribute of a competing alternative *k*. Their expressions are shown below:

$$E_{X_{qji}}^{P_{ji}} = \frac{\partial P_{ji}}{\partial X_{qji}} \cdot \frac{X_{qji}}{P_{ji}} = \beta_{qji} X_{qji} \left(1 - P_{ji}\right), \text{ for direct elasticity} \qquad (17)$$

[14] This fact has some important implications for the identification and specification of choice models. For instance, it is usual that ASCs are included into the model specification, as in (4), because this implies for the error terms a zero mean by construction. Since only differences in utility matter, also only differences in ASCs matter. Then, for a set of *J* available alternatives, simplicity reasons lead to normalize to zero one of the *J* ASCs implying the requirement of only *J-1* ASCs.

[15] For further details about these goodness-of-fit tests, see Louviere et al. (2000).

$$E_{X_{qki}}^{P_{ji}} = \frac{\partial P_{ji}}{\partial X_{qki}} \cdot \frac{X_{qki}}{P_{ji}} = \beta_{qki} X_{qki} P_{ki}\text{, for cross elasticity}^{16} \tag{18}$$

Nevertheless, one of the most important behavioral outputs of the CL model is the possibility of valuing the attributes and alternatives, because it allows knowing the welfare implications of specific policies. In other words, *compensating variation* (CV) can easily be estimated from the CL model. CV is an estimate of welfare change that, according to Morrison et al. (1999), 'shows the change in income that would make an individual indifferent between the initial and subsequent situations given an implied right to the current situation'. In this sense, CV can be described by the next expression:

$$V_0(X_z, X_{q(L1)}, m) = V_0(X_z, X_{q(L2)}\ , m - CV) \tag{19}$$

where, for simplicity reasons, income m is assumed to be the only individual characteristic, $X_{q(L1)}$ and $X_{q(L2)}$ are different levels of attribute q and X_z represent other marketed goods. In this context, the valuation of changes in attribute levels where there are multiple options can be done by applying the expression proposed by Hanemann (1984):

$$CV = -\frac{1}{\hat{\beta}_{COST}} \left\{ \ln\left(\sum_j e^{V_0} \right) - \mathrm{L}n\left(\sum_j e^{V_1} \right) \right\} \tag{20}$$

where V_0 and V_1 represent the utilities of the initial state (i.e. *policy off* context or current situation) and the subsequent stage (i.e. *policy on* context), respectively, and $\hat{\beta}_{COST}$ is the parameter estimate of the monetary attribute (i.e. cost or price of the alternative).[17] When the choice set includes a single before and after option, expression (20) becomes as follows:

$$CV = -\frac{1}{\hat{\beta}_{COST}} (V_0 - V_1) \tag{21}$$

[16] As it can be seen, cross elasticities of an alternative *j* with respect to a variable of an alternative *k* is the same for all $j \neq k$, which is known as *uniform cross elasticities property* and is directly derived from the IID assumption.

[17] The negative of the parameter estimate Of the monetary attribute is interpreted as the *marginal utility of money* or *marginal utility of income*. This is because an increase in cost decreases income, and, hence, the coefficient of cost registers the change in utility associated with a marginal decrease in income.

In some cases, before and after options may differ only because of changes in a single attribute *q*. In this case, expression (21) is reduced to a simpler one. When the functional form of the utility specification is linear both in parameters and attributes, and additive, the marginal value of a change within a single attribute for continuous data can be represented by:

$$CV = -\frac{\hat{\beta}_q}{\hat{\beta}_{COST}} \Delta X_q \tag{22}$$

where $\hat{\beta}_q$ is the coefficient of the *q* attribute and ΔX_q is the quantitative change. However, for qualitative attributes, CV is represented by:

$$CV = -\frac{\hat{\beta}_{q(L2)} - \hat{\beta}_{q(L1)}}{\hat{\beta}_{COST}} \tag{23}$$

where the numerator shows the difference between the coefficients of the attribute levels representing the discrete change. When this change is respect to an opt-out option in the choice set not described by attributes, and, hence, involving the no election of any alternative by individuals, $\hat{\beta}_{q(L1)}$ equals 0 and expression (23) is equivalent to the *marginal rate of substitution* (MRS) between the qualitative attribute and the cost. This is always the case for changes in quantitative attributes.[18] In this sense, MRS becomes a financial indicator of the WTP for one additional unit of the non-monetary attribute, known as *implicit price* or *attribute value*, holding constant all other influences on utility. The possibility to allow the valuation of an attribute to be a function of its levels enriches the point estimates into a distribution of values referred to as *valuation function*. It shows how the attribute value changes as one of their levels changes. In this sense, a positive value of the WTP for one attribute at a specific level means a positive impact of this level on utility of individuals, and, therefore, on probability to choose the specific alternative having this level. However, a negative value supposes individuals asking for being *compensated to accept* the specific attribute level considered. The concept of change at the margin has acquired a high importance within the framework of decision processes. Additional costs and benefits generated by a change from a pre-defined baseline situation represent the main issue in terms of policy making. In this context, valuation techniques based on RUM models

[18] When the monetary attribute is not considered, the ratio of two utility parameters, holding constant all other influences over utility, shows the differences in value attached to different attributes.

and, therefore, capable of measuring welfare, become very appealing methods for economic analysis. When the advantages that SP methods present over RP methods, especially the one of constructing hypothetical scenarios that allows ex-ante analyses, can be combined with welfare outputs from RUM models, as happens in CEs, then the attractiveness of the valuation technique increases.

IV. EXPERIMENTAL DESIGN PROCESS

Experimental design is the basis of any SP experiment. Following the words of Hensher et al. (2005), 'an *experiment* defined in scientific terms involves the observation of the effect upon one variable, a *response variable*, given the manipulation of the levels of one or more other variables'. Statistical design theory has become the tool that makes possible decide which manipulations to make and when to make them in order to obtain parameter estimates in a way that they are not confounded by other factors. The property of design control is viewed as one of the main reasons that explain the growing use of CEs in valuation research over the last years. The possibility to test of certain hypotheses of interest through the control and manipulation of the experiment is argued to be one of the major advantages of the CE.

To outline the basic ideas of the experimental design process, this work follows a similar structure as the one used by Hensher et al. (2005). It is presented in Figure 1, in which the sequential stages required to implement an experimental design are grouped into three broad categories:

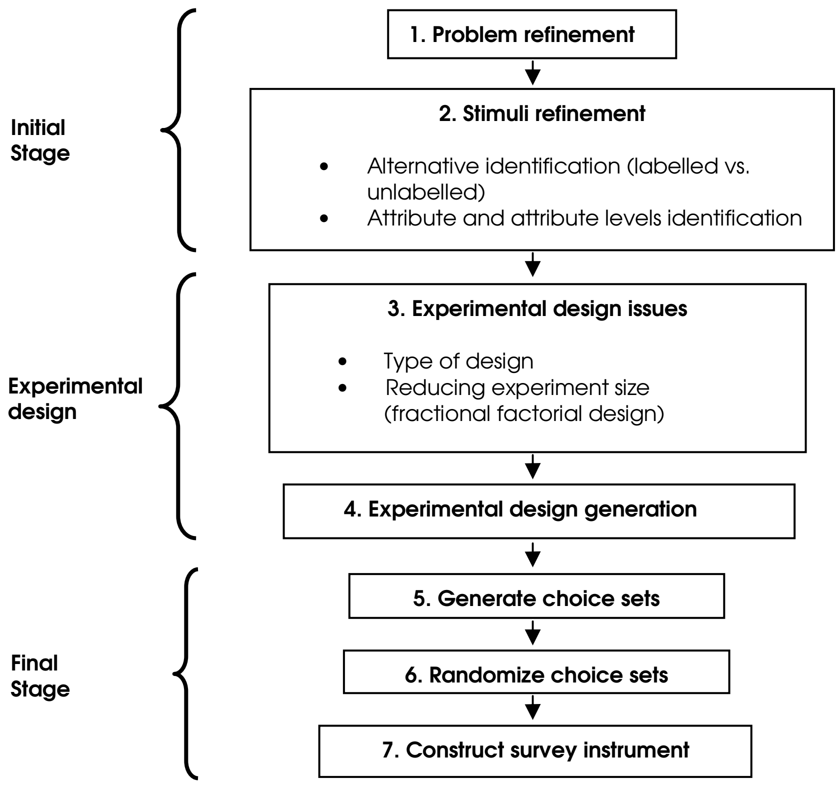

Figure 1. Experimental design process

4.1 Initial stage of the design process

Tasks concerning the initial stage of the experimental design process are two. The first step is the *problem refinement*. In this task, the decision problem is characterized and research problem related hypotheses are generated. The second phase of the initial stage consists of *stimuli refinement*, a task through which alternatives, attributes and attribute levels are identified.

4.1.1 Alternatives identification

In experimental design, the profiles of the alternatives within the choice set are described by *treatment combinations*, that is, the combination of different attributes at different levels or *treatments*.[19] When defining the set of mutually exclusive alternatives, the universal but finite set of the available ones related to the context being studied should be taken into account to be consistent with the *global utility maximizing rule*.[20] However, the complexity associated with a very large number of alternatives has led to the use of strategies aimed at reducing their number. In fact, one of the tasks of analysts in the second part of the initial stage is to cull the alternatives from the universal but finite set of the available ones. They have several ways to do it. First, they can assign to each individual a randomly sampled number of alternatives (plus the chosen one) taken from the global set. This allows considering the entire population of alternatives, but at the cost of complex experimental designs. Second, they can reduce the number by excluding *insignificant* alternatives. However, it is a somewhat subjective process. In this context, one of the most commonly used ways to make the number of alternatives smaller consists of using *unlabelled* (i.e. *uninformative*) alternatives, that is, alternatives with *generic* names that, in contrast with *labelled* ones, convey no information beyond that provided by their attributes.[21]

Experiments that use labelled alternatives are known as *labelled experiments*, whereas those that assign them generic names are called *unlabelled experiments*. A comparison between them shows an array of advantages related to the latter. On one

[19] The terminology in experimental design literature is extensive. Thus, attributes can also be called *factors*, and levels, *attribute levels* or *factor levels*.

[20] This rule states that when an individual acts rationally in choosing an alternative, he acts as if he is maximizing utility, and this says nothing about the choice set, the alternatives and the attributes. Then, this is seen as a *global* assumption that does not permit to exclude any relevant information from an individual perspective.

[21] The use of unlabelled alternatives is a good strategy to be applied in environmental valuation studies, in which the most important goal is to value attributes and not to predict market shares of actual labelled alternatives.

hand, in unlabelled experiments, it is not required the identification and use of all the alternatives within the universal but finite set, because the attribute levels can be so broad that they are useful for expressing features of different types of alternatives. As already said, this helps reducing their number. In addition, only one *generic utility function* to the general class of good or service defined by an alternative is estimated. This implies the calculation of only *generic parameters* and, therefore, the need of a less number of degrees of freedom and, in turn, the use of a smaller design size. In this context, only *within-alternative orthogonality* is of importance. On the other one, the IID condition is less likely to be met under labelled experiments, which favors the use of unlabelled alternatives. This is because labels can be viewed as the levels of an attribute called *label*, and this can lead to correlated alternatives due to the fact that respondents perceptually relate the alternatives to the attributes used within the experiment. Moreover, in labelled experiments respondents can make assumptions regarding omitted attributes in such a way that they assign them different impacts upon utility depending on the labelled alternative considered. This contradicts the IID condition under which omitted attributes, represented by the unobserved component of utility, have the same influence on the choice of each alternative. Furthermore, some of the omitted attributes could be common to two or more alternatives. Then, their inclusion in the random component of the utility could make the alternatives correlated leading to non-zero off-diagonal variances and, therefore, contradicting the IID condition.[22]

4.1.2 Attributes and attribute levels identification

Attributes identified to be used in an experiment can be common or different between alternatives. If they are common, their levels are not required to be the same. In any case, a key issue is to identify attributes in a way that there is no *inter-attribute correlation*, that is, respondents must *cognitively* perceive the attributes as different. If they are perceptually related, although being statistically uncorrelated, respondents could consider the experiment not seriously and final results could be biased.[23] With respect to the attribute levels, it is common to use *attribute levels labels* in order for survey respondents to better understand them. They can be represented in both nominal and ordinal qualitative terms and quantitative terms. Their *extreme ranges* (i.e.

[22] In spite of this, the use of labelled alternatives can be recommendable when, for study purposes, alternative-specific parameters are to be estimated or when realism is required.

[23] Attributes encountered in environmental valuation problems may be highly correlated by natural processes. Then, they are not intrinsically separable. If they are treated as independent in a CE, respondents can be confused and reject the scenario. In these cases, it is recommended to use only feasible combinations of attributes (Holmes & Adamowicz, 2003). In addition, it is to recall that if the experiment follows an orthogonal design criterion, then the use of correlated attributes compromises the results.

end-points) are usually observed values of levels. However, for prediction purposes, they can also be values outside the identified range provided they are credible for respondents. There is no rule to decide the number of levels for each attribute. In this context, what matters is that the more the levels of an attribute, the more the information captured in the utility space derived from a single attribute at varying different levels.

4.2 Experimental design

4.2.1 Experimental design issues

Experimental design considerations constitute one of the most relevant parts of the experimental design process. The issues consider in this part are crucial for the validity and reliability of the results obtained. The success of a CE will highly depend on the efforts done in this stage.

4.2.1.1 Type of design

The number of attributes and the number of levels affect directly the size of the experimental design, which influences the choice of one class of design or another. The most common kind of design is the *full* or *complete factorial design*, where all possible treatment combinations are enumerated. As earlier said, a treatment combination is a combination of different attributes at different levels or treatments. For a choice to take place at least two alternatives are needed. In other words, one of the most important experimental design concepts in CEs is the choice set. The number of attributes and levels determine the number of choice sets that can be constructed to carry out a CE application. In particular, the number of choice sets required in a full factorial design is given by L^{MA}, where L is the number of levels, A the number of attributes and M the number of alternatives. The design is then called L^{MA} *factorial* or L^{MA} *design*.

To make more understandable what a complete factorial design is, an example is presented. For facility reasons, it is assumed the existence of only one alternative, with two attributes, namely A_1 and A_2, each at three levels, L_{A11}, L_{A12}, L_{A13}, and L_{A21}, L_{A22}, L_{A23}, respectively.[24] Under these assumptions, the full factorial design is as presented in Table 1:

[24] For CEs, each alternative in the choice set would be codified in the same way.

Table 1. Full factorial design

Treatment combinations	A_1	A_2
1	L_{A11}	L_{A21}
2	L_{A11}	L_{A22}
3	L_{A11}	L_{A23}
4	L_{A12}	L_{A21}
5	L_{A12}	L_{A22}
6	L_{A12}	L_{A23}
7	L_{A13}	L_{A21}
8	L_{A13}	L_{A22}
9	L_{A13}	L_{A23}

Experimental design literature has also created coding formats to represent the treatment combinations. One of these coding structures is based on assigning a unique number to each attribute level, from *0* to *L-1*. Returning to the example of Table 1, this coding structure for the full factorial design is given by:

Table 2. Full factorial design coding

Treatment combinations	A_1	A_2
1	0	0
2	0	1
3	0	2
4	1	0
5	1	1
6	1	2
7	2	0
8	2	1
9	2	2

However, in accordance with the goal of obtaining parameter estimates not confounded with other factors, most researchers prefer to use *orthogonal coding*. In this case, the values for codes are such that their sum over any given column is equal to *0*. For this purpose to be accomplished, when one level is assigned a positive number, the second level is assigned the same value, but negative. When the number of levels is odd, the code for the median level is *0*. By convention, the levels are assigned the odd numbers *1, -1, 3, -3, 7, -7*, etc. Values *5* and *-5* are not considered. According to this coding criterion, the design of Table 1 becomes as follows:

Table 3. Full factorial orthogonal coding

Treatment combinations	A_1	A_2
1	-1	-1
2	-1	0
3	-1	1
4	0	-1
5	0	0
6	0	1
7	1	-1
8	1	0
9	1	1
Column sum	0	0

In this context, it is easy to see that a large number of attributes and/or levels irremediably leads to a large design size. Indeed, the more attributes and/or levels, the more the treatment combinations obtained and, hence, the more the choice sets. However, a large number of choice sets has been argued to be an important drawback of CEs, because it is commonly related to task complexity and derived fatigue and learning effects, and this can easily bias the experiment results. In this sense, full factorial design becomes useful only when the number of attributes and/or levels is small. If it is large and, therefore, so it is the quantity of choice sets, researchers usually apply different strategies oriented to reducing their number (Carson et al., 1994).

4.2.1.2 Reducing experiment size

A first approach to reduce the design size consists of reducing the number of attribute levels. However, this is at the cost of reducing the amount of information given by the observations. It can be considered a good method if it is believed that there exists a linear relationship between part-worth utilities of the levels. A second approach consists of *blocking the design*. It can be carried out by adding a new attribute that, in terms of experimental design, involves adding a new orthogonal column, whose number of attribute levels serves as indicator of the number of blocks in which the design will be broken down.[25] On the other side, it can also be done by listing the choice sets in random order and then subdividing the list to obtain blocks of *reasonable* size. Whatever the way of proceeding, when using this strategy, the sample size required for

[25] It might not be always possible to add a new design column for blocking without increasing the number of treatment combinations as, for every design, there exists a finite number of orthogonal columns available that the analyst may choose from.

a blocked design increases as the number of choice sets within the block decreases for a fixed number of combinations.[26]

Another option is to use only a fraction of the total number of choice sets, that is, a *fractional factorial design.*[27] Albeit being one of the most commonly used, this strategy presents some problems. In using fractional factorial designs, respondents are presented with subsets of choice sets or rows, so that particular effects of interest can be efficiently and independently estimated. Nevertheless, this does not always ensure orthogonality.[28] An equal number of responses for each row is needed if orthogonality wants to be obtained.[29] In other words, *sampling* becomes crucial to implement *orthogonal fractional factorial designs*. In addition, randomly selecting a number of choice sets without replacement from the total set does not ensure a *statistically efficient fractional factorial design*. For it to be achieved with the minimum number of choice sets, and considering the final goal of *generating the smallest orthogonal design*, the following steps must be carried out:

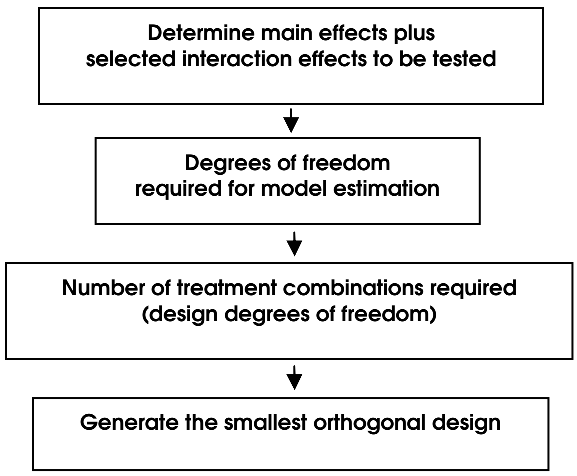

Figure 2. Stages in deriving fractional factorial designs (Hensher et al., 2005)

The first stage in Figure 2 forces analysts to do an in-depth analysis of both *main effects* (MEs) and *interaction effects* (IEs). Confoundment between them has been argued to be a problem in fractional factorial designs. Determination of MEs and IEs has to do with

[26] Blocking strategy requires the assumption of respondents' homogeneity of preferences or, alternatively, a way to deal with respondents' preference heterogeneity.

[27] Reduction of the number of choice sets with which individuals have to be presented could be also achieved by *combining a fractional factorial design with a blocking strategy.*

[28] Full factorial designs mathematically display orthogonality. This means that the columns of the design display zero correlations and, hence, the attributes are statistically independent.

[29] Once orthogonality is achieved for a given number of rows, the removal of columns will not affect it.

model specification. An *effect* is the impact a particular attribute level has on choice. It is the response generated when moving from one level of a given attribute to the next, whilst holding the levels of the all other attributes constant (Garrod & Willis, 1999). In this sense, a ME is the direct independent effect of each attribute upon choice. For experimental designs, an *effect is the difference in treatment means*. Then, a ME is the difference in the means of each level of an attribute (i.e. *marginal means*)[30] and the overall or *grand mean* (i.e. *intercept*), such that the sum of the differences equals zero. In this context, the total number of MEs that can be estimated is equivalent to the number of attributes present in the design. More specifically, the attribute weights are the estimates of MEs. An IE is an effect obtained by combining two or more attributes. Interaction occurs when the preference for the level of one attribute is dependent on the level of another one.[31]

When specifying a model, the number of *degrees of freedom* required for estimation also affects the functional form of utility. Degrees of freedom are the number of observations in a sample minus the number of independent (linear) constraints placed upon it during the modeling process. In deriving fractional factorial designs, knowing this number means knowing the information required for estimation purposes suitable for obtaining the *minimum number of treatment combinations*. The number of degrees of freedom is highly related to the number of parameters of the model. In this context, model specification and, hence, the number of parameters, depends not only on the determination of MEs, plus selected IEs, but also on the consideration of *linear* and/or *non-linear effects*. To understand the importance of choosing between these two latter kinds of effects, let's go to the expression (5), which is additive, linear in parameters, and linear in attributes, and represents a *MEs only specification*:

$$V_{ji} = \beta_{0ji} + \sum_{q=1}^{Q} \beta_{qji} X_{qji} \qquad (5)$$

When linear effects are considered, a coding structure in which levels are assigned numbers from *0* to *L-1* implies that changes in utility for unitary changes in the level of one attribute are linear, that is, constant. Let's consider, for instance, only the attribute

[30] Marginal means represent the marginal effects upon utility of the attribute levels.

[31] It is to recall that *interaction* is not the same as *correlation*, because interaction measures the impact that a combination of attributes has upon choice, whereas correlation shows the relationship between these attributes. In CEs, there are two more types of effects: the *own effects* and the *cross effects*. The first ones are the MEs and/or IEs of an alternative on its own utility or choices, and will be the only statistically significant if the IID condition holds. The second ones refer to the MEs and/or IEs of other alternatives on a particular alternative's utility. The presence of cross effects implies that more cross effects than would be expected by chance should be statistically significant.

X_{1j} [32], supposing it has three levels, *L11*, *L12* and *L13*, codified as *0*, *1* and *2*, respectively. The effect upon utility of each of these levels, other factors being equal, is as follows:

$V_j = \beta_{0j} + (\beta_{1j} x 0) = \beta_{0j}$, the utility associated with the level *L11* (24)

$V_j = \beta_{0j} + (\beta_{1j} x 1) = \beta_{0j} + \beta_{1j}$, the utility associated with the level *L12* (25)

$V_j = \beta_{0j} + (\beta_{1j} x 2) = \beta_{0j} + 2\beta_{1j}$, the utility associated with the level *L13* (26)

It can be observed that when passing from one level to another one of the attribute, utility always changes in an amount equal to β_{1j}. The result would have been the same if levels had been assigned orthogonal codes, that is, if *L11*, *L12* and *L13* had been given the values *-1*, *0* and *1*, respectively. However, this kind of response for attribute level changes is not always what happens in reality. Indeed, when qualitative attributes are used, realism forces to use non-linear effects. In this case, other types of coding approaches are needed. One of them is *dummy coding*. It consists of decomposing the attribute *q* in *L-1 dummy variables*, such that there are *L-1* parameters associated with this attribute. Returning to the example above, it means that for attribute X_{1j}, two dummy variables must be created, namely one for *L13* and one for *L12*, with parameters β_{13j} and β_{12j}, respectively. Automatically, the remaining level *L11* becomes the base level and, hence, its related utility is obtained by assigning value *0* to both dummy variables. Thus, the utility associated with each attribute level is the next:

$V_j = \beta_{0j} + \beta_{13j} x 1 + \beta_{12j} x 0 = \beta_{0j} + \beta_{13j}$, for the level *L13* (27)

$V_j = \beta_{0j} + \beta_{13j} x 0 + \beta_{12j} x 1 = \beta_{0j} + \beta_{12j}$, for the level *L12* (28)

$V_j = \beta_{0j} + \beta_{13j} x 0 + \beta_{12j} x 0 = \beta_{0j}$, for the base level *L11* (29)

According to (27), it can be seen that the ME of level *L13* is equal to β_{13j}, that is, the difference between the marginal mean of the level $\beta_{0j} + \beta_{13j}$ and the grand mean β_{0j}. The same happens for the level *L12*, with a ME equal to β_{12j}. However, it can not be calculated the ME for the base level, because the utility corresponding to it is equal to the *grand mean* β_{0j}, and β_{0j} is the average overall utility level of the utility function. The utility associated with the base level of an attribute and the overall mean are then

[32] For simplicity reasons, subindex *i* has been removed.

perfectly correlated. To avoid this problem, researchers usually use another kind of coding structure. It is called *effects coding*. In this case, *L-1* dummy variables are also created, but the coding for the base level is not *0* but *-1*. Thus, for the example above, the utilities associated with levels L_{13} and L_{12} are the same as in (27) and (28), but the one associated with the base level is as follows:

$$V_j = \beta_{0j} + \beta_{13j} x(-1) + \beta_{12j} x(-1) = \beta_{0j} - (\beta_{13j} + \beta_{12j}) \quad (30)$$

Expression (30) shows that the utility associated with the base level is not anymore confounded with the grand mean β_{0j}. From (30), it can easily be derived the ME for the base level, which is $-\beta_{13j} - \beta_{12j}$ around β_{0j}.

When deciding which type of effects to estimate, one has to take into account that the more complex the part-worth utility function is, the better off is to move to a more complex coding structure capable of estimating more complex relationships. In this sense, it is to recall that the influence of an attribute upon the utility function is better understood when the number of parameters to be estimated is higher. Figure 3 shows that the degree of accuracy to establish the true utility function by analysts is higher when the number of non-linear effects is greater:[33]

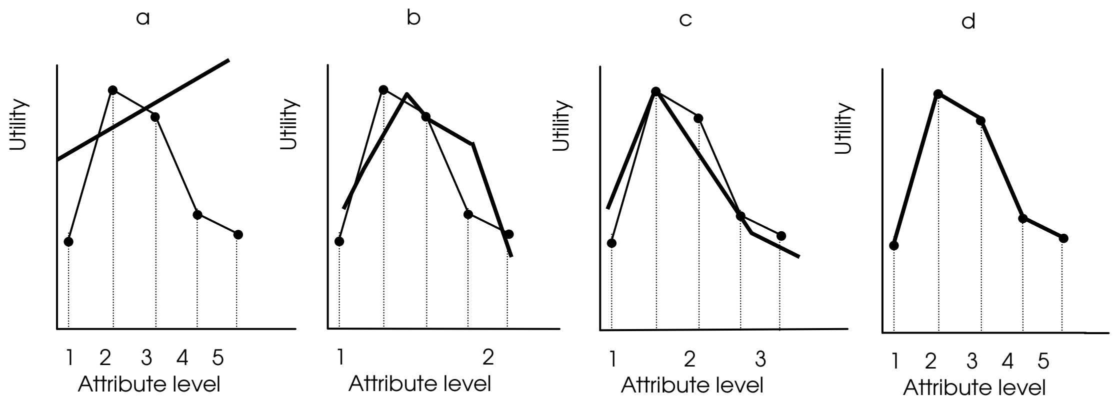

Figure 3. Estimation of linear vs. non-linear effects (Hensher et al., 2005)

In this context, and, as earlier said, especially if qualitative attributes are present in the design, dummy or effects coding is necessary when specifying the model. In addition,

[33] The estimation of a single parameter for an attribute produces a linear estimate (i.e. slope). This corresponds to the case 3a of Figure3. On the other side, an attribute estimated with two dummy or effects parameters is known as a *quadratic estimate*. For subsequent dummy or effects parameters, attributes are known as *polynomials of degree L-1*.

the choice between linear and non-linear effects has more implications rather than those related to model specification. Thus, it has been argued that quantitative attribute levels could directly be introduced into the experimental design in substitution of design or orthogonal codes if only linear effects are considered. This would allow calculating the marginal value for both a unitary and a non-unitary change of the attribute beyond the discrete points used as attribute level labels. However, quantitative levels are required to be equally spaced if correlation wants to be avoided and, then, orthogonality maintained. This is so even though the underlying design or orthogonal codes remain orthogonal.

Given knowledge of all the above, it can be said that by assuming a MEs only model the degrees of freedom necessary for a design depend on the type of effects (linear and/or non-linear) and also on whether the experiment is labelled or unlabelled. Table 4 shows the minimum treatment combinations (i.e. choice sets) requirements for a *MEs only fractional factorial design*:

Table 4. Minimum treatment combinations requirements for MEs only fractional factorial designs[34]

Effects	Experiment	
	Unlabelled	Labelled
Linear	$A+1$	$MA+1$
Non-linear (dummy or effects codes)	$[(L-1)\ x\ A]+1$	$[(L-1)\ x\ MA]+1$

(Hensher et al., 2005)

The smallest possible *MEs plan* is determined by the total degrees of freedom required to estimate all implied MEs, which, at the same time, are determined by summing the separate degrees of freedom in each ME.[35] The more the number of degrees of freedom required for estimation purposes, the larger the design size.

If interaction terms want to be estimated, the design size will be higher. The degrees of freedom for estimating them also depend on the specification of the utility function. Then, if interaction factors come from linear MEs, the number of degrees of freedom will be equal to *1* irrespective of the number of terms in the interaction. On the contrary,

[34] The additive factor 1 refers to the degree of freedom of the random component of the model.

[35] Assuming non-linear effects, each attribute requires *L-1* degrees of freedom for MEs to be estimated.

the degrees of freedom associated with a *q*-order interaction term when non-linear MEs are considered are given by:

$$\left(L_1 - 1\right)\ x \ldots x \left(L_q - 1\right) x \ldots x \left(L_Q - 1\right); \quad q = 1, \ldots, q, \ldots Q \tag{31}$$

where L_q represents the number of levels of attribute *q*. However, it is usual to specify models with only MEs, either treated as linear or non-linear. In other words, for simplicity reasons, it is usually accepted to loose a part of information and, therefore, generate the *smallest ME design*[36]. Nevertheless, it is to recall that for a smallest ME design to be orthogonal, a number of treatment combinations higher than the number of degrees of freedom indicated in Table 4 will be required.

4.2.2 Experimental design generation

The *generation of the experimental design* is carried out through the use of specific software. It is based on generating design columns for all MEs and for the selected IEs by using orthogonal coding. The number of IEs to be tested for is decided prior to design generation. In this sense, if it is believed that two attributes cause an IE, then all two-way IEs columns must be generated. In the next phase, *attributes must be allocated to design columns*. One of the main issues to deal with in this stage is the calculation of the *correlation matrix*. This allows observing the correlations between MEs as well as between MEs and IEs. This is crucial to allocate the attributes to design columns, because they must be allocated to those columns that are unconfounded.[37] If analysts decide that non-linear effects are to be estimated, then attribute levels must be effects or dummy coded in order to calculate then the correlation matrix for the dummy or effects coded design. Effects or dummy coding is done after design columns for each of the attributes have been orthogonal coded. Unfortunately, the introduction of dummy or effects codes leads to correlations within the design and, therefore, provokes a loss of design orthogonality.

[36] By only considering MEs, it is supposed that IEs are not statistically significant. If, in reality, they are significant, then results obtained from the estimated model will be sub-optimal. For the purpose of realism, the analyst can also generate designs that allow for some IEs. In fact, it has been said that a good design strategy should be to use designs that allow estimation of (at least) all two-way interactions whenever possible, because MEs and two-way interactions account for virtually all the reliable explained variance (Louviere et al., 2000).

[37] If, for instance, it is supposed that an IE is caused by the interaction of two attributes for only two alternatives, then four design columns will be required. In case there is more than one interaction design column unconfounded with all MEs, only two of them must be chosen. The attributes causing the IEs must be assigned to the MEs design columns generating the chosen interaction columns. It is also important that the two chosen IEs columns are not confounded with each other.

4.3 Final stage of the design process

The *generation of choice sets* must be done by using understandable attribute levels labels. Pictures can be used as aid. *Randomizing choice sets* is an optional stage aimed at testing for *order effect bias* at the time of model estimation. This kind of bias is related to the effect upon the answers of respondents of possible learning and fatigue effects whilst carrying out the CE. Randomization consists of showing each decision maker within a block (provided the CE has been blocked) with the same choice sets but in different order to one another. However, there not exists a rule about how many randomized versions are optimal and to how many respondents those versions should be physically distributed.

The *construction of the survey* ends the process of experimental design. In this latter phase, the issue of the *choice context* must be carefully managed. In SP applications, researchers must give respondents detailed information about the context in order for them to make their choices meaningfully. It is also important that people consider each choice decision as independent of the other ones. In other words, analysts must care about the fact that the hypothetical scenario presented in each choice set is not compared by individuals with the other ones. The presence of a *do-nothing option* is another issue to take into account. There are two main forms of do-nothing options. One of them is that of *no-choice* or *no-purchase* option. It refers to the choice not to select one of the available alternatives. In environmental economics, for valuing recreational activities this may mean *stay at home* or *do some other activity*. The other one is the *SQ or current alternative*. It can be used if the current RP experience of individuals, invariable across the choice sets, wants to be included into the CE and the new alternatives pivoted from it.[38] Within the environmental valuation framework, for valuing environmental policies, it can mean that individuals are forced to live with current environmental conditions. The inclusion of do-nothing options makes both choices more realistic and welfare measures obtained consistent with demand theory and utility maximization. This is because it avoids forcing respondents to choose one of the *new* alternatives presented.[39] In addition, this kind of alternatives serve as a useful reference point in terms of utility and attribute levels for respondents (Blamey et al., 2001). However, it is to recall that the type of do-nothing option to be considered is of

[38] In this case, quantitative attribute levels are usually described as percentages of the reported levels. This leads to the measurement of the impacts upon choice of percentage changes (not absolute) in attribute levels. Within the framework of do-nothing options, there also exists the possibility of considering a *delay-alternative*, which involves the choice to delay the decision for the present.

[39] In environmental contexts, most applications tend to use three alternatives within the choice sets, including an opt-out option (Banzhaf et al., 2001).

paramount importance, because it can lead respondents to evaluate choice sets in different ways and this can impact results substantially.

At this point, and for the purpose of this work, the main issues underlying experimental design process have been outlined. However, other design considerations remain that must be taken into account in doing an SP experiment. Issues such as the level of aggregation in the analysis, the presence and treatment of preference heterogeneity, or the need for multiple versions of the survey and the number of respondents to be assigned to each version should also influence the design chosen (Carson et al., 1994). Nowadays, one of the main design concerns has to do with efficiency considerations. In this sense, new design criterions based on *optimal designs* have emerged, that is, designs that maximize the amount of information that can be obtained from them. They are used as alternative to orthogonal designs. There are still few studies applying them. Whilst in orthogonal fractional factorial designs attributes are statistically independent, although being statistically inefficient, in optimal designs they are statistically efficient but likely to be correlated. However, although they are aimed at gaining statistical efficiency, this aim is not always achieved. This is because an important efficiency issue has not been given the deserved attention. It is respondents' efficiency.

It is of great importance to consider that total efficiency of a CE is given not only by design efficiency but also by respondent efficiency. As Louviere & Hensher (2000) state, 'humans interact with CEs in ways not considered by the choice modelling community, such that one must take into account not only design efficiency but also respondent efficiency to determine the total efficiency of a CE'. However, in attempting to understand preferences and separate them from noise (i.e. unexplained variation), researchers have used a high array of techniques, such as statistical design theory, econometric specifications or the combination of RP and SP methods, but giving little attention to the choice environments or task demands. The number of choice sets, attributes, levels, the correlation structure of the attributes, among others, are design issues that constitute the choice environments and can highly influence individuals' choices by imposing on them more or less cognitive burden demands (Swait & Adamowicz, 1996). In other words, researchers have paid little attention to task complexity and, hence, to its effect on respondents' efficiency due to possible learning and fatigue effects, which supposes a risk for the reliability of CE results. In this context, although efforts oriented to gaining statistical efficiency in optimal designs studies have been made, efficiency problems can remain. This is the case of a study done by Huber & Zwerina (1996), who propose to use *utility balance experiments* to gain statistical efficiency. Choice sets are balance in utility, that is, they have similar choice probabilities. Then, individuals are asked to choose among options that are close in utility. However, this makes the choice more difficult. The authors show that the gains from statistical efficiency are offset by a decrease in respondent efficiency due to a major task complexity.

Then, using more complex techniques to achieve increases in design efficiency can not be always a good option if respondent efficiency is not given a deserved, careful attention. A major design complexity, though interesting and useful for specific purposes, must not be an issue per se. Following the words of Louviere et al. (2000), 'complex models that *demonstrate* one's statistical and/or mathematical superiority are not *better* models. Rather, better models come from real understanding of the behavior of interest and its antecedent links, which leads to significant insights into behavior before parameter estimation'. For this reason, it is believed that an in-depth understanding of the outlined, and most widely used, design issues of CEs, based on orthogonal designs, can serve as a good basis for the line of research proposed in this work.

V. LITERATURE REVIEW

The first CE application involving environmental issues emerges in the context of the pooled models. This approach, based on the inverse relationship between scale and random component variability, involve the combination of CEs with RP methods to estimate joint models.[40] In this way, advantages of CEs over RP methods can be used to improve the estimates derived from the latter. It is argued that, among other things, joint models lead to an increase in the amount of information available, the possibility of modelling non previously existing scenarios, or goods with attribute levels outside the range of current ones, and the reduction of collinearity offered by the SP statistical designs. The pioneering effort in combining SP and RP choice data is attributed to Ben-Akiva & Morikawa (1990). Some years later, Adamowicz et al. (1994) adopt the approach to infer recreationalists' preferences for alternative flow scenarios for the Highwood and Little Bow rivers in Alberta, Canada. It represents not only one of the first applications of joint SP and RP methods in non-market valuation, together with Ben-Akiva & Morikawa's work, but also the first CE application to environmental management problems.

Pooled model applications involving CEs are usually focused on recreational site choice. The interest in knowing the value of different characteristics of recreational sites through the use of CEs is usually motivated by the existence of conflicts between the users' activities and the environmental features of the sites. To implement public policies oriented to achieving an efficient use of the sites, knowledge is required about the value that users assign to the different site's features. In this way, the effect on use values that policies can generate, and, hence, on users' behavior, can be known. Modelling recreational demand requires eliciting information on users' actual choice. In this context, the most commonly used RP technique to be combined with CEs is the travel cost method (TCM), where individuals' choice is explained as a function of the travel distance and the quality of site attributes. For the joint estimation to be possible, the CE is required to explain the choice of one alternative over the others as a function of the same attributes used in the TCM, where price attribute is proxied by travel distance to the site. The fact that both models reflect the same process of choosing recreational sites, based on the same trade-offs between attributes, makes possible the joint analysis.

[40] Swait & Louviere (1993) show how to estimate the ratio of scale parameters for two different data sets, which can then be used to compare different models or to pool data from different sources.

In this sense, Adamowicz et al. (1994) construct alternatives described by the same attributes both in the CE and in the TCM. However, the choice sets considered in each method differ, because the number of generic alternatives in the CE is generated by an experimental design process, whereas the trip options in the TCM correspond to real sites. For the CE design, the authors distinguish between two kinds of alternatives, standing water alternatives and running water alternatives. Both of them share 8 common attributes, being *fish size*, *water quality*, *swimming*, and *beach*, at 2 levels, and *terrain*, *fish catch rate*, *facilities*, and *distance*, at 4 levels. For the design, they also define alternative-specific attributes, in the sense that the levels to be combined differ depending on whether the alternative is standing or running water related. These specific features are *water feature* and *fish species*, each at 2 levels, and *boating* at 4 levels, for standing water alternatives, and *water feature*, at 2 levels, and *fish species*, at 4 levels, for running water alternatives. The attributes of these alternatives are treated as the collective factorial $(2^{6}\cdot 4^{5})\cdot(2^{5}\cdot 4^{5})$, from which an orthogonal MEs design is selected. The final design uses 64 choice sets and is blocked into 4 sets of 16 choice sets. Each choice set shows one standing water alternative, one running water alternative, and an opt-out option meaning *stay at home* or *any other non-water related activity*. The paper done by Adamowicz et al. (1994) is one of the few CE applications that consider a high number of attributes to construct alternatives. As earlier said, a high number of characteristics is likely to impose important cognitive burden demands on respondents, which can compromise the success of the CE exercise. However, the issue of task complexity is not studied in the paper.

Estimation issues are also important for the reliability and validity of the results. Some authors center their attention on testing different hypotheses with the objective to find the best model specification. This usually forces them to work with more complex statistical models. However, the majority of CE applications concerning joint models use basic model specifications. In this sense, Adamowicz et al. (1994) estimate a CL model for each of the two types of alternatives, standing and running water, finding out, as a main result, that attributes such as *water quality* and *fish catch* are significant determinants of trip destination.

Another important CE application combining RP and SP data is the one carried out by Adamowicz et al. (1997), also based on recreational site choice. The purpose of the paper is to show the role that perceived measures of attributes, obtained by asking individuals to quantify their perceptions about them, play in welfare estimates in comparison with objective measures of the same attributes. The basis of this application can be found in the CE exercise done by Boxall et al. (1996) about recreational moose hunters in Alberta. Then, the number of attributes and levels, and the general features of the experimental design used, which are explained later, are exactly the same. However, in Adamowicz et al. (1997), the authors focus on three different data structures for examining choice behavior when different types of measures of attributes are used. The first one is characterized by the use of choice data generated by revealed choices, a choice set defined by the researcher, and objective measures for attribute data. The second involves choice data generated by revealed choices, a choice set defined by respondents, and the use of perceived measures for attribute data. Finally, as a third data structure, they use choice data generated by stated choices, a choice set defined by the researcher, and the use of constructed attribute data. Models estimations are done with the use of CLs. The main conclusion of the paper is that the joint RP-SP model based on perceptions moderately outperforms the other models. This can be explained by the fact that the variation in the attributes in the perceptions data may capture more of the variation in the observed component of the RUM model rather than the error term, whereas the lack of variation in the objective data may lead to the higher error variance in the RP objective data. However, the authors alert to the fact that if subjective and objective measures are not strongly correlated, then estimation results and welfares measures can be different.

In spite of recognition of the advantages of combining SP and RP methods, there are still few studies in environmental economics doing joint analyses. Recreational site choice has been usually treated in the valuation literature applying CEs outside the framework of pooled models. In fact, as said above, the study done by Adamowicz et al. (1997) has its origins in the more simple CE application carried out by Boxall et al. (1996), in which the authors do an empirical exercise of CVM, based on contingent behavior (CB), and CE to assess the impact of alternative logging methods on recreational hunting values. Although the study area involves 15 Wildlife Management Units (WMUs), the quality change examined is the improvement in moose populations as a result of careful forest harvesting in only one specific WMU. This is because the CVM only permits examining one change in one WMU, whereas the CE allows estimation of welfare impacts due to a change in the levels of any attribute at any of the 15 sites. Then, to compare the CVM and the CE, the latter has to be restricted at the specific WMU used in the CVM. Another requirement to compare both models in case that the CVM is based on CB is that cost attribute must be measured in the same way. Then, it is proxied by travel distance in the CE, as happens in joint models.

For the experimental design, Boxall et al. (1996) use 2 attributes at 2 levels, *forestry activity* and *road quality*, and 4 attributes at 4 levels, *moose population*, *hunter* *congestion*, *hunter access*, and *distance to site*. From the universe of combinations of $(4^4 \cdot 2^2) \cdot (4^4 \cdot 2^2)$, they finally use 32 choice sets, which are then blocked into 2 sets of 16. For each choice set, respondents face three alternatives, two of them being competing WMUs and the third one an opt-out option meaning *not to go moose hunting at all*. This 32 choice sets represent the smallest orthogonal MEs design. On the other side, for model estimation they use a binary logit (BL) for the CVM and a CL for the CE. Results show that when the CE is restricted to the specific site considered in the CVM, that is, when substitutes are not considered, results from the two methods become more similar. Then, the importance of substitutes is highlighted in the paper. It is also emphasized that CEs are attractive for environmental valuation, because they rely on the same model structures as referendum CVM models and discrete TC models. This feature has served as a basis for a lot of applications comparing CE results with values derived from other SP and RP methods as a way to test convergent validity of the CE.

Some other studies concerning recreational site choice make a comparison between the CE and an RP method. This is the case of an application carried out by Hanley et al. (2002) based on modelling the recreational activity of climbing in Scotland. As said earlier, comparisons require using the same attributes in both methods to make the comparison possible. Then, 6 attributes of climbs are used, which are *length of the climb*, *approach time*, *overall quality of the climb*, and *scenic quality*, each at 4 levels; *crowding on the climb*, at 2 levels; and *distance*, as a proxy for cost, at 6 levels. For the experimental design, choice sets are produced using a fractional factorial design. Climbers, presented with either 4 or 8 choice tasks, are asked to choose between two routes described in terms of their attributes and an opt-out option meaning *stay at home*. As a novelty regarding previous studies, the authors are concerned about the role that experimental design issues can play on final results. Then, they do a test for task complexity and a test for rationality and conclude that design decisions seem to have a small impact on WTP and that the majority of respondents behave rationally in answering choice questions.

For model estimation in the CE, Hanley et al. (2002) first use a basic CL model, that is, a model without interactions between attributes and SDCs. After observing IIA violation, they estimate a nested logit (NL) model, with and without interactions with SDCs, that is, both a basic and an extended NL. To be able to compare results derived from the two models, they also construct an extension of the basic CL model by including SDCs. Estimation results show that the inclusion of individual-specific covariates gives similar coefficients for both CL and NL models with respect to their basic case. Then, they conclude that models are robust when including covariates, which is a claim in favor of extended models. However, when comparing parameter estimates from the CL and the NL, Hanley et al. (2002) show that they are larger in the NL. Nevertheless, they state that, in terms of policy-making, IIA violation does not appear to have a great impact on estimates of implicit prices. On the other side, when applying a TCM as a test of convergent validity of the CE, they show that both methods show similar pictures of climbers' preferences over different sites. Then, according to the authors, it is difficult to say which method is better to model recreation demand, because although CE present advantages in terms of non-collinearity and the possibility to use more levels than the current ones, RP methods do never present hypothetical effects.

However, most of CE studies focusing on recreational site choice are centered on only CE applications, without comparing CE results with other methods. In this sense, it can be found the study done by Bullock et al. (1998) about stalking (i.e. deer hunting) as an important land use and activity in Scotland Highlands. They are interested in determining the value that stalkers of red deer place upon different characteristics of their stalking trip and the value of alternative packages of such attributes. The attributes considered to construct stalking trip options are the *cost*, at 9 levels, and 4 physical attributes, *deer number*, *quality*, *activities*, and *landscape*, at 3 levels. The complete design is expressed by the authors as $2^2 \cdot 3^7$, which gives 8.748 choice sets. However, they use a fraction of the full factorial design, which, unlike previous CE studies, is very large. Specifically, they use one-third of the full factorial, divided into sets of 6 choice tasks. Finally, the authors randomize the orders of presentation of the choice sets, which show respondents only two trip alternatives to be chosen, trip A and trip B.

In Bullock et al. (1998), individuals' choice is first analyzed by using a BL model. Then, by asking respondents if they prefer the characteristics of either of the two trips of the choice sets to the same characteristics of their last trip, a SQ alternative (i.e. last trip) is introduced in the estimation, which is carried out by using a multinomial logit model. The final stage of the analysis is based on indirectly obtaining respondents' rankings of the alternatives without directly asking individuals to rank them. This is done by simply asking for the preference between A or B in case they have answered 'A and B' from a set of four possible answers (A and B, just A, just B, or neither) to the first choice set question. The authors measure the welfare change for different constructed stalking packages and conclude that there is a way to satisfy both deer managers and conservationists.

Morey et al. (2002) carry out an application where the potential of the CE for modelling recreation demand is again examined. In a context where the growing popularity of mountain biking in many areas in the US has led to increased trail degradation and conflicts among users on single track, they apply a CE to see the effects that trail characteristics and access fees have on trail selection in an attempt to estimate the benefits and costs of trail closure and access fees to users. The attributes considered are *total length of trail*, *percentage of trail that is single track*, *total vertical feet of climbing*, *number of peaks along trail profile*, and *entrance fee*, each at 3 levels, and the one indicating if the *site is used by hikers/equestrians*, at 2 levels. 50 pair-wise choice sets are constructed by randomly pairing sites from the 36 realistic sites identified by the focus groups, and replacing any pairings which displayed dominance. The 50 choice sets are blocked into 10 sets, creating 10 different versions of the survey, each with 5 pairs. Surprisingly, neither a SQ option nor an opt-out option is included as a alternative in the choice sets, which, according to the authors, makes impossible the assessment of consumer surplus (CS) per year and only allows calculating CS per ride, because no information is obtained about the desired frequency of rides given the chosen site. The CL estimated assumes that budget affects site choice. Therefore, and unlike other studies, it is used an income-effect model, in which income enters the utility function non-linearly. As a main result, Morey et al. (2002) observe that CS estimates vary across bikers in terms of household budget, gender and interest in mountain biking, amounts that depend on the number of substitute sites and the trail characteristics and fees, if any, at those sites. As an additional contribution of the paper, a BT is simulated to show how the model and parameter estimates can be transferred to estimate the benefits and costs to mountain bikers in a specific area.

At this point, the most known CE applications centered on modelling recreational demand have been outlined. In general terms, researchers have achieved positive results when using CEs to value the effect of site environmental improvements on use values. These applications show that CEs can successfully compete with RP methods when the study object is focused on use values. However, as said in section II, the growing use of CEs in environmental valuation since the 1990s has been motivated by the numerous advantages it has over other SP methods, especially over the widely used CVM. Implicitly, this means that CEs have been mainly used to infer values that only SP methods can estimate, that is, passive use values. In this sense, most of authors have centered their efforts on explaining choices in different environmental quality settings defined as results from specific environmental management programs and mainly described by non-use value characteristics.

The firs CE application estimating passive use values is carried out by Adamowicz et al. (1998a). It is based on estimating non-use values for a threatened woodland caribou management program in Alberta, Canada. Most of studies valuing conservation programs concern comparisons between the CVM and the CE as a way of doing convergent validity tests that state the capability of CEs to measure passive use values. Then, the authors compare results from a CVM, involving, in this case, a dichotomous choice (DC) format, and a CE to test for differences in preferences and error variances arising from the two methods. For both techniques to be compared, they describe the improvement to be analyzed in the same way in the CVM and in the CE. For the construction of the woodland caribou preservation alternatives, they use 5 attributes, each at 4 levels, which are *mountain caribou population*, *wilderness area*, *recreation restrictions*, *forest industry direct employment*, and *annual changes to provincial income tax*. Levels are varied above and below the current ones to examination of both WTP and WTA (i.e. willingness-to-accept) for attribute changes. It can be seen that the authors consider one socioeconomic indicator, *forest industry direct employment*, as a possible determinant of utility gained by the management program. This trend is followed by some further studies that consider individuals assign non-use values to specific socioeconomic attributes. However, not always these features are statistically significant, as happens in this exercise. Scenarios are constructed from a $4^{5} \cdot 4^{5} \cdot 2$ orthogonal MEs design. Finally, 32 choice sets are used, blocked into 4 versions of the questionnaire with 8 choice scenarios presented to each respondent. Choice sets present two management alternatives and a SQ option.

Adamowicz et al. (1998a) estimate a CL model, both in a linear and a non-linear form, for the CE, the CVM, and also a pooled model, and present the results of the quadratic specifications. When comparing the CE and the joint model, they observe there is a large increase in the welfare measure when passing from their linear to their quadratic forms. The authors attribute this result to the nonlinearity of preferences over caribou. It is suspected that this nonlinearity is due to the background information given to respondents, which states that current caribou population level is smaller than the one of the viable population (i.e. small risk of extinction). Then, it is not surprising to see that the marginal utility of caribou declines dramatically after the viable population level is reached. Stating the role of background information represents one of the main contributions of the paper. Regarding the comparison between the CVM and the CE, they find that error variances are not significantly different between both methods. Then, it is concluded that the CE is a good for measuring non-use values.

Hanley et al. (1998a) attempt to estimate the wildlife and landscape benefits associated with the Environmentally Sensitive Areas (ESAs) scheme in Scotland by applying a CE and a CVM. To construct generic ESA alternatives for the CE, they consider the attributes commonly affected by the ESA management provisions. Then, *broadleaved woodland*, *moorland*, *wetland*, *dry stone dykes*, and *archeological sites* are used, each at 2 levels, one corresponding to the authors' predictions of ESA landscapes with ESA management agreements, and another one corresponding to the no ESA management agreement. The cost attribute is assigned 8 levels. An orthogonal MEs design is constructed, creating pair-wise comparisons, which gives a possible $2^5 \cdot 2^5$ design size. According to focus groups considerations that respondents can cope with up to 8 choice pairs each, the final sample size becomes 256 persons. In each choice pair, respondents are asked to choose between two ESA management alternatives and a SQ option. However, they also have the possibility not to choose by answering *they don't know which option to choose*.

For the CE model estimation, Hanley et al. (1998a) use both a linear and a quadratic CL model, observing that the latter performs better. Results show positive marginal values for all the attributes. When comparing the results from the CVM and the CE, they observe that the overall WTP for ESA policies are higher for the CE than for the CVM. However, taking into account that it is not always the case that results from the CVM are higher than the ones from the CE, they state that how to choose characteristics from the very large set available and how this choice impacts on the total package welfare measures becomes an important issues because it can highly influence final results. Then, they recommend that the best option is to use a CVM when the objective is to value some overall policy package or environmental resource.

Colombo et al. (2006) also find that WTP values from the CVM when substitutes are considered are smaller than the ones from the CE. They compare the value of soil erosion control programs from both methods, as well as estimate the values for different attributes of soil conservation plans using the CE. In particular, they try to identify people's preferences towards reducing the off-farm effects of soil erosion in the Alto Genil watershed, in Andalusia. Because the main off-site impacts of soil erosion are water quality, desertification of the landscape and loss of wildlife habitat, they use 5 attributes at 3 levels, *landscape change* (i.e. desertification of the semiarid areas), *surface and ground water quality*, *flora and fauna quality*, *rise of agricultural productivity* (in number of jobs created), *area of project execution in Km*^{2}, and the cost attribute *extra tax*, at 6. The levels describe the likely future conditions with and without the implementation of soil erosion reduction projects. The set of attributes and levels forms a universe of 1,062,153 possible combinations. By means of experimental design techniques, an orthogonal fraction of the complete factorial is drawn, representing the smallest orthogonal and balanced design. It yields 108 combinations to be presented to respondents, which are then blocked into 27 groups. Each respondent faces 4 choice sets, and is asked to choose between a SQ option, which represents the expected environmental situation in the watershed in 50 years if no soil conservation measures are implemented, and two more alternatives showing expected situations in 50 years with soil policies implemented.

Colombo et al. (2006) estimate three models for the CE. Firstly, they estimate a basic CL model that only includes attributes. Then, in order to capture heterogeneity, they estimate an extended CL model with SDCs and attitudinal variables, observing that it performs better and passes the Hausman & McFadden test (1984). However, as another way to incorporate heterogeneity, the authors decide to estimate a random parameter logit (RPL) model. As happens in other studies, results show that allowing for heterogeneous preferences makes little difference to welfare estimates from the extended CL model and the RPL model. One of the main conclusions of the work is that the welfare estimates obtained in the study span the current subsidy that the Andalusia's Government gives to farmers that adopt soil conservation measures.

However, other studies find similar results when comparing the CVM and the CE. It is the case of Jin et al. (2006), who carry out the exercise in the context of developing countries. They attempt to know which methods for measuring contributions to well-being are more appropriate as well as how institutions can assure that economic values are reflected in private and public choices. Specific features of developing countries, which tend to have a complex political, institutional, cultural and socioeconomic background, forces to deal with environmental problems in different ways. Then, information is required about their societies' preferences. Jin et al. (2006) compare the results of a double-bounded DC-CVM and a CE in a valuation of solid waste management programs in Macao, a special administrative region in China. The authors state that solid waste incineration has been given a top priority over other waste disposal methods in Macao due to the small geographic area and the high cost of land. However, capacity for incineration is expected to be insufficient, which forces to identify efficient waste reduction strategies, including waste segregation and recycling. To construct the waste minimization alternatives, 4 attributes are considered, which are defined as *waste segregation and recycling at source*, *waste collection frequency*, *noise reduction in waste collection and transportation process*, each at 2 levels, and *monthly garbage fee per person charged*, at 4 levels. The study involves a $2^3 \cdot 4$ MEs factorial experimental design. However, the design is based on minimal overlap and utility balance principles, which leads to finally use 24 options, blocked into 8 choice sets. In each choice set, respondents are asked to choose between a SQ situation with no costs but more environmental pressure, and a new program with costs. They also have the possibility of choosing 'none' when they do not like any of these two alternatives. For the CE, two CL models are estimated, a basic one, considering only attributes as determinants of individuals' choice, and a second one, extended with the addition of SDCs and attitudinal variables. They observe that the extended model performs better, giving positive marginal values for all the attributes. When restricting the CE to the same improvement offered in the CVM, they find there is no significant difference between the estimated values of changes in solid waste management programs derived from the two methods.

Hanley et al. (1998b) carry out a study in which estimate the external benefits of possible changes in landscape elements in public forests due to changes in management. Unlike other studies, they use an open-ended format for the CVM. The authors think that, although the open-ended format is not based on RUM models and, hence, makes impossible to treat the CVM and the CE as theoretically equivalent, this comparison can still serve as a good convergent validity test. For the design, only 3 attributes, *shape of the edges*, *type of felling*, and *species mix*, each at 2 levels, are used. A MEs design presents respondents with 4 choice tasks, in which they are asked to either choose between two forest designs or the SQ option. For the CE model estimation, it is used a mixed logit (ML), both in a linear and a quadratic form. As a main conclusion, they observe that the latter performs better and use it to state that marginal WTP for all the attributes are positive. Another interesting contribution of the paper is that preferences for users and non-users are different, which suggests the use of different experimental designs for each kind of individual.

In a more recent CE exercise, Christie et al. (2006) try to identify problems surrounding the economic valuation on changes on biodiversity on UK farmland, especially those related to people's limited understanding of complex environmental goods. In particular, the study areas are Cambridgeshire, with a predominantly intensively arable area that supports low levels of biodiversity, and Northumberland, with high levels of biodiversity and lower intensity of land use. They report the results from a CVM on 3 biodiversity enhancing policies, which are biodiversity enhancement related to agri-environmental schemes, the one related to the re-creation of wildlife habitats, and biodiversity loss from farmland associated with development activities, and from a CE that examines the value of biodiversity attributes. They also examine, through a series of valuation workshops, the effect of information deficit, which typifies the knowledge level of most members of the general public regarding biodiversity. To construct alternatives representing policies on biodiversity conservation and enhancement on farmland, they consider 2 ecological attributes, *habitat quality* and *ecosystem processes*, and 2 more within the anthropocentric framework, *rare, unfamiliar species of wildlife*, and *familiar species of wildlife*. Each attribute has 3 levels, except the cost feature, which represents the annual increase in taxation over the next five years and has 5 levels. From the universe of $3^4 \cdot 5^1$ alternatives, a fractional factorial design is generated, which is then blocked to assign the options to 10 bundles of 5 choice sets per respondent. In each choice set, individuals choose between three alternatives, a SQ representing a declining in biodiversity and two improvement options. The results from the CVM show that people place positive values on increases in biodiversity. However, when estimating a CL model for each study area in the CE, evidence from the results suggests that the public support policies that target *rare unfamiliar species of wildlife*, being this evidence less clear for *common familiar species*. Then, as a main conclusion, the authors state that people care about biodiversity but not in how it is achieved. Results from the workshops approaches show that information exchange and group discussion help to reduce the variability of value estimates.

Within the framework of CE applications involving valuation of environmental programs, some authors compare the CE with other SP techniques different from the CVM. In this sense, it stands out an exercise done by Riera & Mogas (2004) focused on a comparison of marginal WTP estimations using a contingent ranking (CR) and a CE. They authors estimate the mean WTP of a given population for changes in their welfare due to a variation in the quantity or quality of some of the attributes that Catalonian forests provide. They define the attributes according to some of the most typical forest functions, such as recreational activities, CO2 sequestration, and soil erosion prevention when increasing in a given amount the surface of forests in Catalonia. Then, 6 attributes are considered, defined as *picnic*, *driving*, *mushrooms*, each at 2 levels, and *CO2*, *erosion*, and *price*, each at 4 levels. Because the authors compare results from two methods belonging to the conjoint analysis approach, experimental design procedures must be applied to both techniques. In this sense, there exist $(2^3 \cdot 4^3) \cdot (2^3 \cdot 4^3)$ possible combinations of afforestation alternatives in the CE, and only $2^3 \cdot 4^3$ in the CR. To select the number of alternatives to be presented to respondents in the CE, an orthogonal fractional factorial design is applied. In this way, 64 sets of pairwise comparisons are obtained, which are grouped into 16 versions of 4 choice decisions. Individuals must choose between a SQ option of no afforestation and two afforestation alternatives. In the CR, 16 alternatives are obtained after having applied an orthogonal fractional factorial design, which are then grouped into sets of 4 alternatives. Then, respondents are asked to rank the 4 alternatives in their order of preference, including the SQ option.

One of the major contributions of that Riera & Mogas (2004)'s work is to obtain CR and CE results from separate samples, unlike other studies (Adamowicz et al., 1998a; Boxall et al., 1996; Hanley et al., 1998b). For the estimation, they use a CL model and a rank-ordered logit model and observe that the marginal WTP for the different methods are different. In particular, the values from the CE are higher than the ones from the CR. Significance of the different WTPs is tested and it is shown that equivalence is rejected. To see if differences persist when data for the different methods are derived from a single sample rather than collected from separate samples, it is derived a pairwise choice dataset from the data obtained in the CR exercise, that is, a CE is simulated. Results show that the simulated CE is more efficient than the CR. The new CE estimates are lower than the ones from the original CE but higher than the ones from CR. The main result, however, is that there are significant differences between the estimates of the original CE and the simulated one, whereas the estimates of the simulated CE are not significantly different from those from the CR. Then, the authors conclude that the differences between CE and CR are due to the different method employed and not due to the different experimental design used.

Willis et al. (2002) also compare a CE and a CR. They carry out an application in a context in which the development of new sources of supply by water companies to ensure that supply and demand for water are kept in balance can conflict with local wildlife interests. They use a case study of a potential water resource development in south-east England. In particular, it is presented the appraisal of a project to artificially recharge an aquifer with river water during the winter period, and abstract the water in the summer during drought conditions. In this context, the authors want to estimate people's preferences for the possible environmental impacts of this project development. The trade-offs investigated are those between increased security of water supply against environmental changes. Willis et al. (2002) consider 2 attributes related to criteria for increased security of supply, *frequency of hosepipe bans*, at 3 levels, and *risk of water supply interruptions*, at 4, and 2 more related to criteria for environmental impacts, *changes in bird and plant diversity*, and *increase or decrease in river levels*, both at 3 levels. The cost attribute represents the change in household water bill and has 4 levels. Respondents are given 4 choice cards, in which a SQ and two project implemented alternatives are presented. When estimating the models, they conclude that the CE is the better because both the log-likelihood and the Akaike information criterion indicate that it is the model closest to the true situation. Results show that people are willing to trade-off security of supply against environmental protection and also to pay towards ensuring that the environment is protected through the implementation of the project. However, the authors attribute this result to the fact that current water supply is very secure and, hence, current levels of supply interruption are negligible.

Other authors valuing preservation programs compare the CE with RP methods. An example is found in the work done by Scarpa et al. (2003), who center their application in the context of developing countries. They state that the conservation and correct assessment of existing biodiversity of plants and animals employed in agriculture is very important for sustainable development. In this context, the management of animal genetic resources requires many decisions that would be easier to make if information on the economic value of populations, traits and processes were known. For this reason, they do a CE survey designed to elicit traders' preferences for various cattle traits, because this species provides a large contribution to many underdeveloped regions. The paper focuses on Maasai Zebu breed as a first crude proxy for the gene pool found within that indigenous breed. Then, as a test of convergent validity, they also do a hedonic pricing approach based on actual observed market transactions at the same time and in the same markets as the CE to see the effect of breed on market prices. For the design, they consider 5 attributes, s*ex*, *breed*, *body condition*, and *price of the animal in Kenyan Shilling*, all at 2 levels, and *slaughter weight*, which is the estimated slaughter weight in Kg. Respondents must choose between two hypothetical cattle purchase choices and an opt-out option. Each of them faces 8 choice tasks. The CE is estimated through a CL model, a mixed logit model to account for unobserved heterogeneity, and a panel version of mixed logit models to account for dependence between the sequential choices made by the same respondent. When comparing results from the CE and the RP method, it is observed that the CE produces estimates on marginal values similar to the ones obtained by the theoretically more valid method of hedonic regression. This allows stating that the CE is a good method for estimating cattle traits relevant in market transactions for Maasai traders. Accounting for taste and variance heterogeneity does not change this conclusion.

In spite of all these studies comparing CEs with other methods, it is to say that an important part of the CE applications focused on valuing preservation programs only carry out simple CE exercises. In addition, most of them have a common denominator based on the consideration of socioeconomic factors as determinants of the utility gained from environmental management policies. In this sense, Morrison et al. (1999) define a passive use value associated with job losses as the value of *preventing job losses.* The setting for their application is the Macquarie Marshes, a major wetland in New South Wales, Australia. The high number of environmental values provided by marshes, such as provision of an important habitat for waterbirds, filtration that improves downstream water quality, and provision of high-quality stock feed, helps to define the 5 attributes that describe the wetland management alternatives, which are *water rates*, *wetlands area*, *frequency of waterbird breeding*, and *endangered and protected species present*, at 4 levels, and *irrigation-related employment*, at 3 levels. Experimental design procedures following a MEs orthogonal design lead to 5 sets of possible options for each respondent. Each choice task shows different options available for the management of the Macquarie Marshes, including a SQ, in which the size of Marshes is expected to decline, and two improvements options. However, individuals have the possibility to not choose any of these three options, which automatically means they prefer more water allocated for irrigation and, hence, a decrease in water to the wetlands. Estimates are obtained through the estimation of two CL models, a basic one, showing only the weight of attributes in choices, and an extended one that includes both SDCs and attitudinal variables in addition to the attributes. It is observed that random taste variations lead to IIA violation in the basic model. Then, the extended one becomes better and does not violate the IIA assumption. Marginal attribute values are positive and significant, being especially higher for the attribute *frequency of waterbird breeding*. Another contribution of the paper is the inclusion of a dummy variable that captures whether respondent is intended to visit the marshes in the future or not, a variable that results significant and positive. In this way, this paper becomes one of the first CE applications estimating option values.

Wetlands are among the Earth's most productive ecosystems providing a high array of ecological functions and services, which translate directly into economic functions and services, such as flood protection, water supply, improved water quality, commercial and recreational fishing and hunting, and the mitigation of global climate change. However, an increasing anthropogenic pressure is compromising their health. Then, it is not surprising that a lot of CE studies involving the valuation of environmental and socioeconomic attributes are centered totally or partially on wetlands management programs. In this sense, it stands out the study carried out by Mallawaarachchi et al. (2001) aimed at the assessment of the WTP for the protection of areas of natural vegetation in Herbert River District of north Queensland, where wetlands and natural woodlands may be cleared to grow sugarcane. Their aim is to identify best practicable land-use options that maximize regional profits and minimize environmental externalities in land allocation. They use as a socioeconomic factor future income in the region. Regional income is used as a proxy of income and employment effects associated with sugar industry activities. For the construction of the alternatives, 4 attributes, each at 3 levels, are used. Thus, they use *annual levy on land rates*, *income in region in 2005*, *teatree woodlands in 2005*, and *vegetation along rivers and in wetlands in 2005*. An orthogonal experimental design is used to assign attribute levels to options. From the 3^4 full factorial design, a fraction is used to allocate the 81 choice configurations to 9 blocks of 9 choice sets, over 9 versions of the questionnaire. Respondents are asked to choose between a SQ option, whose levels are different from the ones used for the design, and fixed, and two incentive scheme options.

Mallawaarachchi et al. (2001) estimate a NL model, in which the branch-choice equation consists of choosing between *doing something* or *doing nothing*. At the second level of the nest, respondents choose between *doing something* options. As a result, Herbert residents, who benefit significantly from the sugar industry, are willing to pay for environmental protection. In particular, their preference for wetland preservation is much higher than the one for teatree woodlands, which is consistent with reality where the available area of wetlands has declined faster than the area of teatree woodlands. Then, the estimates reflect this relative scarcity value. Furthermore, the regular recreational use of wetlands by residents and their growing conservation motives for riparian areas may also have contributed to this assessment. The authors also demonstrate that environmental values of wetlands are comparable to returns from commercial production of sugar cane and that the values of teatree woodlands are comparable to returns from commercial grazing, which allows them to state that the CE is a good tool for estimating the trade-offs between direct financial and environmental impacts in development activities. In addition, and following Morrison et al. (1999), they also construct a dummy variable called *visit* to capture option values.

Othman et al. (2004) have also been interested in estimating non-use values related to socioeconomic and environmental features within the wetlands valuation framework. Their study is based on the estimation of different resource management options of the Matang Mangrove Wetlands in Perak State, Malaysia, where mangroves have been gazetted as a protected forest since the 1920s. They estimate an employment value, because they think that those not directly affected by the businesses might derive non-use values from local employment opportunities provided to the local people by the various commercial activities undertaken in the wetlands. For the design, they use 5 attributes. These are *environmental forest area*, *visitation rates*, *number of migratory bird species*, *number employed*, and the *contribution to a trust fund*, all of them at 3 levels. After having applied a MEs fractional factorial design, respondents are asked to choose between three options in a set of five choice tasks. The three alternatives are the SQ and two options involving maintaining or increasing the environmental forest area while maintaining or decreasing the production forest area. The authors estimate a basic CL model and an extended one including SDCs and attitudinal variables.

However, the IIA test of Hausman & McFadden (1984) reveals violation of the assumption for both models. For this reason, they estimate a NL model to capture heterogeneity in a better way. Nevertheless, they observe that implicit prices estimated from the extended CL and the NL model do not differ substantially, which means that heterogeneity in respondents' preferences has little effect on attribute values, a result that is common in other CE studies.

Birol et al. (2006b) also consider socioeconomic features in a study aimed at assessing the benefits generated by the sustainable management of the Cheimaditida wetland, in Greece. The authors use 5 attributes, 3 of them at 2 levels, *biodiversity*, *open water surface area*, and *research and education*, and 2 attributes, *number of farmers retrained in environmentally friendly employment*, and a *one-off payment a wetland management fund*, at 4. It is used an orthogonal MEs design from which 32 pair-wise comparisons of alternative wetland management scenarios are constructed. These are randomly blocked into 4 different versions, each with 8 choice sets. Each set has two wetland management scenarios and an opt-out option. For model estimation, Birol et al. (2006b) use a basic CL model, only with attributes, in which they observe IIA condition is not violated. However, because preferences tend to be heterogeneous, they also estimate a RPL model but only considering the attributes as the determinants of utility, that is, accounting only for unobserved (i.e. random) heterogeneity. Results support choice specific unconditional heterogeneity. However, to be able to explain the sources of this heterogeneity, that is, to explain conditional (i.e. observed) heterogeneity, they extend the RPL model by including interactions with SDCs and attitudinal factors. The authors also enrich the paper by estimating a latent class model as another way for accounting for preference heterogeneity. They conclude that there is heterogeneity across the public, and, on average, people derive positive and significant values from sustainable management of this wetland.

There are other studies not focused on wetlands that also attribute importance to the non-use value of socioeconomic attributes. In this sense, Rolfe et al. (2000) carry out a study centered on the estimation of the non-use values that Australians might hold for the preservation of rainforests in Vanuatu, a Pacific Island. Rainforests are recognized throughout the world for their biological richness and ecological importance. However, tropical deforestation brings about diverse environmental problems such as impacts on climate, loss of plant and animals species, and impacts of ecosystem loss, and short-term and long-term production problems. According to that, the chosen attributes to form management alternatives are a mixture of environmental and socio-economic features. Within the environmental attributes, it can be found *location*, at 7 levels, *area*, *rarity*, and *special features*, all 3 at 3 levels. The socioeconomic ones are *potential to visit*, and *the effect on local (indigenous) people*, both at 3 levels, and the *cost* at 4. Unlike the work done by Morrison et al. (1999) and the one carried out by Mallawaarachchi et al. (2001), the authors try to infer an option value by defining an attribute representing the kind of possible visits to the rainforest, if any, in the future (*potential to visit*). The 7 attributes are combined in $8^1 \cdot 4^6$ ways to form 32.768 possible different profiles of rainforest protection options. An experimental design process is used to select the sets of profiles that are presented to respondents. For each choice set, individuals have to choose only between two protection options, although they are given the possibility of answer *Not support either option*. Rolfe et al. (2000) estimate a basic CL model. However, IIA tests indicate that the model does not fully conform to the underlying IID conditions. Then, for improving model fit and removing IIA violations, SDCs are included. They observe as an important result of the expanded model that the probability of choice can be largely predicted according to attributes and SDCs, and that non-observed variables (represented by constant term) are insignificant. The main results show the importance of socioeconomic attributes in the overall assessment of preservation proposals.

On the other side, CE studies can be found that do not consider socioeconomic attributes as determinants of the value of the environmental resource to be managed. Some of them are also concern about wetlands policies. This is the case of the study carried out by Carlsson et al. (2003), who estimate WTP for different characteristics of a wetland area in Staffanstorp, southern Sweden. For the design, they use 7 attributes, which are *total cost*, at 4 levels, *biodiversity*, at 3 levels, and *surrounding vegetation*, *fish*, *fenced waterline*, *crayfish*, and *walking facilities*, all at 2 levels. In this study, choice sets are created using a linear D-optimal design procedure, which gives 60 choice sets that are then blocked into 15 versions, each containing 4 choice sets. Respondents must choose between three alternatives, the third one being the SQ with no improvement. The authors use both a CL and a RPL model to estimate the coefficients. Results show the RPL is superior to the CL, that is, there is heterogeneity of preferences for the attributes. The robustness in the RPL results is caused by the advantages characterizing the RPL models, which are the fact that the alternatives are not independent (i.e. the model does not exhibit the IIA property) and there is an explicit account for unobserved heterogeneity. However, the gain in terms of precision of the WTP estimates is unclear. As a test of internal validity, they also test for stability, by comparing the estimated preferences for two different orders (one in which half of the respondents receives the choice sets in the order 1, 2, 3, 4, and the other half in 4, 3, 2, 1), and conclude that hypothesis of stable preferences cannot be rejected.

Other CE applications have their origins in the Water Framework Directive (WFD)'s concerns. It constitutes a major regulatory reform of water resources management within the EU and is aimed at achieving good ecological status in all European waters. In this sense, Hanley et al. (2006) estimate the values people place on improvements in some indicators of ecological status that ordinary people see as important but differ from the ones considered by ecologists. The CE is done in the context of an improvement to the ecology of the River Wear, in County Durham, England, and in the River Clyde, in Central Scotland. The attributes used are 3 river quality features, *in-stream ecology*, *aesthetics/appearance*, and *bankside conditions*, each at 2 levels (good and fair, being fair consistent with current conditions on the rivers, and good consistent with regulators' expectations as to what will likely constitute good ecological quality status under the WFD), and *water rates*, at 5 levels. It is used a fractional factorial design, not blocked due to the simple nature of the design. Then, each respondent faces 8 choice questions, and each choice set has three alternatives, two alternative catchment management plans for each river, and a SQ option.

Hanley et al. (2006) estimate a basic CL for the pooled sample and for each river. However, after having observed that IIA is violated, RPL models are estimated. General results show that people value water quality improvements. Nevertheless, they observe that the parameter estimates for River Wear are very similar, whereas the ones for River Clyde show larger differences in attribute values. On the other side, in calculating the implicit prices for the sample of River Wear, they state they are very similar in the CL and the RPL, whereas for River Clyde, they are not significant for CL but they are for RPL. Thus, they conclude that heterogeneity must only be accounted for in River Clyde. The paper also carries out BTs to see what errors are likely to be experienced if BT procedures are used as part of implementing the WFD. This is done because it is think that in implementing the WFD, BT methods will be needed due to the high costs of valuation studies. However, they find that equality of parameters between the two rivers is rejected and conclude that BT is, therefore, not advisable.

Focusing on the CE capability to evaluate alternative policy options and to overcome some of the CVM disadvantages, Blamey et al. (1999) carry out an application that consists of a consumer-based assessment of five possible future water supply options for a future Australian Capital Territory (ACT) population in the vicinity of 450,000, with particular attention to environmental costs. In particular, they want to know community rankings for these options. To construct the choice sets, they use generic alternatives by considering 6 attributes, all of them at 3 levels, defined as *reduction in household water use*, *use of recycled water*, *increase in water charges*, *improvement in river flows*, *number of species with habitat loss*, and *color of grass in urban areas*. It is used an orthogonal full 3^6 factorial design. However, to reduce the number of alternatives, a one-twenty-seventh fraction is used, from which 27 combinations are obtained that are assigned to 3 blocks, such that any one respondent is confronted with no more than 9 different options in 9 choice sets. Individuals must choose between two options showing changes in water supply and a base option without policy supply. The novelty of this design is that the base option is not a SQ alternative, but a policy option, because without water policy supply, water demand restrictions are strictly necessary. Then, WTP involving movements from this option are conditional on the increase in the household water cost having to be made in case no water supply is carried out. This implies that, in reality, it is not attempted to estimate CS for changes from water supply conditions. The authors present results for three model specifications. First, they estimate a basic CL model, with only an attribute specification. Then, they add SDCs in an additive form, interacted with ASC. Finally, they estimate a third model that includes interactions of SDCs with selected attributes. To obtain a ranking of the set of feasible options, the authors calculate the choice probabilities, that is, the probability of the average individual choosing an option when the only other alternative in the choice set is the base option. After this, they calculate the probabilities corresponding to an expanded set involving all five management options in order to obtain market share estimates involving the proportion of ACT residents favoring each option.

CE studies concern about transitional economies. In a context in which the possible demise of traditional farms has been cited as one of the costs of EU accession, economic transition and development, with the possible subsequent loss in food or livelihood security of a lot of people, valuation studies are required to know society's preferences for traditional life styles to draw conclusions about the sustainability of agrobiodiversity. In this sense, Birol et al. (2006c) carry out a study oriented to estimate the private economic value that farm households derive from four components of agrobiodiversity in home gardens (i.e. Hungarian traditional farms). In particular, they focus on three ESAs, Dévaványa, Őrség-Vend, and Szatmár-Bereg. The attributes used for the options presenting agrobiodiversity managed in home gardens are 5. They are *cultivation of landraces*, *traditional method of integrating crop and livestock production*, and *use of organic production practices*, all at 2 levels, and *crop variety diversity* and *cost*, this latter measured as self-sufficiency in terms of percentage of annual household food consumption that it is expected the home garden will supply, both at 4 levels. Orthogonalisation procedure is used to recover only MEs, consisting of 32 pair-wise comparisons of home garden profiles, randomly blocked to 6 different versions, 2 with 6 choice sets, and the remaining 4 with 5 choice sets. Each choice set shows two home garden alternatives and an opt-out option.

The authors estimate CL models with logarithmic and linear specifications for the pooled sample (the three areas), although they only present estimates for the linear form. After having observed that IIA condition is violated, they estimate a RPL model. However, as other CE studies show, this has a little effect on implicit prices. For the CL, all attribute values are negative. They also present results for each of the areas, by estimating a CL model where observed heterogeneity is accounted for, that is, by interacting SDCs with choice-attributes. The main result is that home garden attributes contribute positively and significantly to the utility of farmers in areas of Hungary, and the relative importance depends on the social and economic characteristics of farm families and their location.

Other authors have moved their attention towards the role of specific features on environmental conservation, as is the case of McGonagle & Swallow (2005). They examine the role of public access in WTP for coastal land conservation by residents from Rhode Island. The authors argue that some people may desire additional land conservation specifically to facilitate public access to natural resources, such as coastal areas, whereas other individuals may favor programs directed at ecological quality goals. These interactions could interact to mislead estimation of the relative value of land conservation. Then, if provision of public access is a key factor in voter or donor support for conservation, failure to provide access may undermine recent initiatives. As attributes to construct the alternatives, they use 6 physical features, *shore type*, and *water type*, both at 4 levels, *location*, and *development level nearby*, both at 3 levels, *unique scenic quality*, and *unique ecological quality*, both at 2 levels; 3 management attributes, *access level*, at 4 levels, *facilities proposed*, at 3 levels; *law enforcement*, at 2; and *cost*, at 5 levels. It is used a fractional factorial MEs design with 64 choice sets. Respondents must evaluate attributes for two parcels of coastal land which are hypothetically available for preservation or to forego preservation of both (i.e. an opt-out option). The authors estimate a basic CL model, only with attributes, and an extended one including SDCs, and observe there is heterogeneity in preferences regarding the role of public access. More specifically, results show that public access is very influential on respondents' WTP to support conservation, and, while some individuals may identify public access as a good, others may see public access as a bad, or as a conflicting use of sites valued for ecological conservation. This reveals opportunities to optimize conservation programs that serve heterogeneous populations.

At this point, it has been presented a general overview of the most important CE applications concerning environmental issues that have been carried out since the adoption of the method by environmental economists in the early 1990s. This literature review has outlined the natural resources on which researchers have mainly focused over the last years, as well as the experimental design procedures and econometric analyses mostly used in CEs.[41] By reviewing the literature, the increasing role that CEs are playing in other emerging subdisciplines based on inferring individuals' preferences for non-market values other than the environmental ones has also been stated.[42]

The fact that the number of CE publications in the 2000 decade is superior to the number of published studies in the 1990s shows the growing acceptance that the method is gaining among the economists concerned about the assessment of non-market values, especially the environmental ones. However, the technique is in its infancy and there is still a lot of work to do. Diverse environmental issues remain to be analyzed in the context of CEs. Others topics remain even to be examined in the whole economic valuation framework. One of them wants to be the study object of the PhD thesis whose proposal is presented in this work. It is thought it represents one of the most challenging research topics in the context of economic valuation and that the CE can play an important role to deal with it. The need for this specific research line is explained in the next section.

[41] A summary of the most common features of CE applications carried out since the 1990s is shown in table I in the annex.

[42] This is, for instance, the case of a study done by Itaoka et al. (2006), who estimate the WTP for reduction of mortality risks caused by fossil fuel power generation versus mortality risks caused by nuclear power generation; or the application done by Mazzanti (2003), who applies a CE to analyze visitor preferences and estimate their WTP for incremental changes in services associated with the stock of cultural heritage.

VI. A FUTURE RESEARCH LINE

The majority of CE studies are focused on valuing different environmental management programs rather than modeling recreational demand. Then, it seems to be there is a major interest in using the CE from a non-use value perspective by taking advantage from its capabilities as SP method. In this sense, a lot of studies are centered on comparisons between the CVM and the CE in an attempt to state the advantages of the latter over the former. They usually show concern about issues related to woodlands, wildlife, landscape, and biodiversity, among others. Another important part of CE applications involving valuation of preservation programs only carries out CE exercises and, although the study objects are also diverse, the most valued one is represented by wetlands.

However, an in-depth analysis of these applications shows that no much attention has been paid to the valuation of environmental costs, especially in terms of ecological discontinuities and possible irreversibilities. Curiously, in a context of an increasing anthropogenic pressure over the ecosystems, very little is said about their carrying capacity, and attribute levels representing thresholds of environmental sustainability are considered in very few studies. In this sense, it stands out the work done by Adamowicz et al. (1998a), who use the level of *viable population* for caribou woodlands, after which marginal utility of caribou declines dramatically. Unfortunately, however, the majority of CE studies are not constructed on ecological and economic integration issues. In words of Adamowicz (2004), 'there has been no much success in measuring passive use values and ecosystem service values', and, hence, 'this area presents the most significant challenge for environmental economists, in such a way that efforts will necessarily include consideration of sustainability and irreversibilites as well as the complexities of ecosystem-social systems interactions'.

In this context, the effect that ecological thresholds and potential irreversibilities can have on individuals' utility is overlooked. However, it is expected that changes in the environmental state of natural resources, likely accompanied by changes in the supply of goods and services provided by them, affect people's preferences. In other words, it is expected that ecological non-linearities cause non-linearities in the valuation function. In this sense, an SP method as the CE can play an important role to measure values, because, on one hand, SP methods do not need people to have made choices in response to thresholds effects in the past, which is in line with the unpredictable thresholds effects occurrence, and, on the other one, CEs can be designed to value a variety of plausible ecosystem scenarios so that it can be described the sensitivity of the obtained values to each possible outcome.

This project proposal is within the framework of the EC-funded *Thresholds of Environmental Sustainability* project. Then, as explained in section I, it wants to be focused on the environmental degradation of coastal waters due to eutrophication processes. The high inversely correlation between the loss of water transparency and the change in water color, as two of the consequences of the proliferation of algal blooms, and environmental aesthetics makes water recreational values interesting values to be assessed in a context of ecological discontinuities, because aesthetics is supposed to highly influence them. Then, the analysis of the effects of ecological thresholds on individuals' utility can be treated through the implementation of a CE involving alternative water recreational scenarios, in which the levels of one of the attributes, defined in terms of water transparency or water color, serve as the indicators of different water environmental states. Non-linear effects of the ecological attribute on the valuation function can be then stated if the marginal values for its different levels are found to be statistically significant and different, that is, the hypothesis of equivalence between the marginal values of the levels is rejected. According to that, a first research question to be answered can be defined as:

Do ecological thresholds effects cause non-linearities in individuals' valuation function?

However, to deal with this research topic, a specific amount of scientific information about the levels of ecological attributes representing different environmental states of the resource is needed. Then, work must be done under ecological and economics issues. In addition, for the thresholds effects to be well captured, a non-linear specification of the utility function is required. In this context, it is expected that ecological discontinuities lead to estimate complex relationships, which leads to the use of a more complex part-worth utility function. Put in other way, it is required a utility specification with a high number of non-linear effects capturing the weight of each level of the ecological attribute. In this sense, it is to recall that the majority of CE applications use a number of levels for environmental attributes that ranges from 2 to 4. However, this range can be not enough to capture the complex relationships that are expected to be associated with thresholds effects. In fact, two levels can only represent a linear relationship. Then, a minimum of three levels would be required to capture some threshold effects. This leads to consider a higher number of levels for the ecological attribute. It is thought that the analysis of the marginal values for each ecological attribute level when a different number of levels is considered is an essential task to deal with ecological non-linearities. In other words, to draw conclusions about the best way to capture the influence of thresholds effects on individuals' utility, the implementation of different CEs only differing from the number of levels for the ecological attribute is required. Nevertheless, carrying out a valuation study is expensive. Furthermore, in CEs, a higher number of levels leads to more treatment combinations and, hence, to more choice sets, which is directly related to a higher sample size. That is, CE becomes doubly expensive. In this context, the role that simulation techniques can play is very important.

In a context of ecological thresholds, another interesting related topic is to test for heterogeneity. Most of CE studies, especially the ones focused on the non-use value approach, test for heterogeneity (Birol *et al.*, 2006b; Birol *et al.*, 2006c; Carlsson *et al.*, 2003; Colombo *et al.*, 2006; Hanley *et al.*, 2006; McGonagle & Swallow, 2005; Othman *et al.*, 2004; Scarpa *et al.*, 2003). By analyzing heterogeneity in the context of ecological non-linearities, it can be known if different users of the resource respond similarly when facing ecological thresholds. Different non-linearities in the valuation function would help to know which the most affected users are, if any, in the context analyzed. In a coastal water framework, it is expected that bathers are the most affected water users when water quality is degraded. However, it needs to be demonstrated. In any case, considering heterogeneity issues favors more equitable policy making decisions, because each group of users is treated according to their benefits and costs. All this leads to the second research question:

Do ecological thresholds effects cause the same non-linearities, if any, in different users' valuation functions?

The analysis of thresholds effects, however, is not exempt from problems. The number of levels is also one of the determinants of the choice environment or task demand. It has been shown that complex choice environments highly influence individuals' choices and, hence, taste parameters (De Palma et al., 1994; Heiner, 1983; Mazzotta & Opaluch, 1995; Payne et al., 1988; Simonson & Tversky, 1992). Swait & Adamowicz (1996) show that complexity impacts variances and draw two interesting conclusions from this. First, they find a U-shaped relationship between the variances of the latent utilities and the level of design complexity. Following their words, 'as individuals face increasing complexity, they will respond with increasing information about their trade-offs (i.e. decreasing variance), but beyond some point of complexity, greater inconsistency across individuals will be found, and so variance increases'. Second, they also find a convex function of cumulative cognitive burden for the variance of utility.

This means that, 'a common sequence of events for a respondent in an SP choice task may be learning for some number of replications, followed by the replication of the learned behavior during another number of replications, and finally, fatigue sets, leading to less consistent choice behavior'.

Nevertheless, in spite of the relevance of this topic, few papers concern about it (Bullock et al., 1998; Hanley et al., 2002). Then, although the majority of CE applications uses a number of levels for environmental attributes that ranges from 2 to 4, it remains to be demonstrated if for these experimental designs there is task complexity effects on individuals' choices. For that to be done, a measure of complexity related to the number of levels could be calculated to analyze its relationship with the variance of latent utilities and see if it follows the path described by Swait & Adamowicz (1996). In this sense, it is considered necessary to test for the existence and magnitude of task complexity effects for different CEs whose only difference is their number of levels for the ecological attribute. Trying to capture thresholds effects in the best way by increasing the number of levels for the ecological attribute can not be well accomplished if important task complexity effects emerge. Equilibrium between these two issues must be found. This leads to a third research question:

Is there a U-shaped relationship between the latent utilities for environmental states involving thresholds effects and the level of the design complexity measured as a function of the number of the levels for the ecological attribute?

At this point, it has been outlined the research line that wants to guide the PhD thesis. In particular, it wants to be constructed on the three research questions explained above. The objective is complex, but challenging. Valuing environmental costs for different states characterized by uncertain thresholds effects becomes a crucial task, because it can serve as a preventing tool in case an environmental damage has not happened yet, or as a way to design the preferred restoration project in case the damage has happened and is reversible. Then, it is thought that important contributions to the valuation literature can be done, especially because the main objective is to cover the existing gap in terms of the valuation of ecological non-linearities. As Deacon et al. (1998) state, 'the most valuable future contributions are likely to emerge from research programs that identify specific gaps or inconsistencies in the current state of the art, and develop empirical or theoretical strategies that will close them'.

REFERENCES

Adamowicz, V., & Boxall, P. (2001). *Future directions of stated choice methods for environment valuation.* Paper presented at the Choice experiments: a new approach to environmental valuation, London, England.

Adamowicz, W. (2004). What 's it worth? An examination of historical trends and future directions in environmental valuation. *The Australian Journal of Agricultural and Resource Economics, 48*(3), 419-443.

Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. J. (1998a). Stated preference approaches for measuring passive use values: choice experiment and contingent valuation. *American Journal of Agricultural Economics, 80*, 64-75.

Adamowicz, W., Louviere, J., & Swait, J. (1998b). *Introduction to attribute-based stated choice methods.* Final report to NOAA Resource Valuation Branch, Damage Assessment Centre.

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management, 26*, 271-292.

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., & Williams, M. (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. *Journal of Environmental Economics and Management, 32*, 65-84.

Alpízar, F., Carlsson, F., & Martinsson, P. (2001). Using choice experiments for non-market valuation. *Economic Issues, 8*(1), 83-110.

Banzhaf, M. R., Johnson, F. R., & Mathews, K. E. (2001). Opt-out alternatives and anglers' stated preferences. In J. Bennett & R. Blamey (Eds.), *The choice modelling approach to environmental valuation* (pp. 157-177). Cheltenham, UK: Edward Elgar.

Bateman, I. J., Carson, R. T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Özdemiroglu, E., Pearce, D. W., Sugden, R., & Swanson, J. (2002). *Economic valuation with stated preference techniques: a manual.* Cheltenham, UK: Edward Elgar.

Batsell, R. R., & Louviere, J. J. (1991). Experimental analysis of choice. *Marketing Letters, 2*(3), 199-214.

Ben-Akiva, M., & Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A, 24*(6), 485-495.

Bennett, J., & Adamowicz, V. (2001). Some fundamentals of environmental choice modelling. In J. Bennett & R. Blamey (Eds.), *The choice modelling approach to environmental valuation* (pp. 37-69). Cheltenham, UK: Edward Elgar.

Birol, E., Karousakis, K., & Koundouri, P. (2006a). Using economic valuation techniques to inform water resources management: a survey and critical appraisal of available techniques and an application. *Science of the Total Environment, 365*, 105-122.

Birol, E., Karousakis, K., & Koundouri, P. (2006b). Using a choice experiment to account for preference heterogeneity in wetland attributes: The case of Cheimaditida wetland in Greece. *Ecological Economics, 60*, 145-156.

Birol, E., Smale, M., & Gyovai, A. (2006c). Using a choice experiment to estimate farmers' valuation of agrobiodiversity on Hungarian small farms. *Environmental & Resource Economics, 34*(4), 439-469.

Blamey, R., Gordon, J., & Chapman, R. (1999). Choice modelling: assessing the environmental values of water supply options. *The Australian Journal of Agricultural and Resource Economics, 43*(3), 337-357.

Blamey, R., Louviere, J. J., & Bennett, J. (2001). Choice set design. In J. Bennett & R. Blamey (Eds.), *The choice modelling approach to environmental valuation* (pp. 133-156). Cheltenham, UK: Edward Elgar.

Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M., & Louviere, J. J. (1996). A comparison of stated preference methods for environmental valuation. *Ecological Economics, 18*, 243-253.

Bullock, C. H., Elston, D. A., & Chalmers, N. A. (1998). An application of economic choice experiments to a traditional land use-deer hunting and landscape change in the Scottish Highlands. *Journal of Environmental Management, 52*, 335-351.

Carlsson, F., Frykblom, P., & Liljenstolpe, C. (2003). Valuing wetland attributes: an application of choice experiments. *Ecological Economics, 47*, 95-103.

Carson, R., Louviere, J., Anderson, D., Arabie, P., Bunch, D., Hensher, D., Johnson, R., Kuhfeld, W., Steinberg, D., Swait, J., Timmermans, H., & Wiley, J. (1994). Experimental analysis of choice. *Marketing Letters, 5*(4), 351-368.

Carson, R. T. (1998). Analysis. Valuation of tropical rainforests: philosophical and practical issues in the use of contingent valuation. *Ecological Economics, 24*, 15-29.

Ciriacy-Wantrup, S. V. (1947). Capital returns from soil-conservation practices. *Journal of Farm Economics, 29*, 1181-1196.

Colombo, S., Calatrava-Requena, J., & Hanley, N. (2006). Analysing the social benefits of soil conservation measures using stated preference methods. *Ecological Economics, 58*, 850-861.

Christie, M., Hanley, N., Warren, J., Murphy, K., Wright, R., & Hyde, T. (2006). Valuing the diversity of biodiversity. *Ecological Economics, 58*(2), 304-317.

Davis, R. (1963). *The value of outdoor recreation: an economic study of the Maine woods.* Harvard University, Harvard.

De Palma, A., Myers, G. M., & Papageorgiou, Y. Y. (1994). Rationale choice under an imperfect ability to choose. *American Economic Review, 84*(3), 419-440.

Deacon, R. T., Brookshire, D. S., Fisher, A. C., Kneese, A. V., Kolstad, C. D., Scrogin, D., Smith, V. K., Ward, M., & Wilen, J. (1998). Research trends and opportunities in environmental and natural resource economics. *Environmental and Resource Economics, 11*(3-4), 383-397.

Diamond, P. A., & Hausman, J. A. (1994). Contingent valuation: is some number better than no number? *Journal of Economic Perspective, 8*(4), 45-64.

Freeman, A. M. (1982). *Air and water pollution control: a benefit-cost assessment*. New York: John Wiley and Sons.

Garrod, G., & Willis, K. G. (1999). *Economic valuation of the environment: methods and case studies*. Cheltenham, UK: Edward Elgar.

Hanemann, W. M. (1984). *Applied welfare analysis with quantitative response models*. Working paper no. 241. University of California, Berkeley.

Hanemann, W. M., & Kanninen, B. (1999). The statistical analysis of discrete-response CV data. In I. J. Bateman & K. G. Willis (Eds.), *Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EC, and developing countries* (pp. 302-441). New York: Oxford University Press.

Hanley, N., MacMillan, D., Wright, R. E., Bullock, C., Simpson, I., Parsisson, D., & Crabtree, B. (1998a). Contingent valuation versus choice experiments: Estimating the benefits of environmentally sensitive areas in Scotland. *Journal of Agricultural Economics, 49*(1), 1-15.

Hanley, N., Mourato, S., & Wright, R. E. (2001). Choice modelling approaches: a superior alternative for environmental valuation? *Journal of Economic Surveys, 15*(3), 435-462.

Hanley, N., Ryan, M., & Wright, R. (2003). Estimating the monetary value of health care: lessons from environmental economics. *Health Economics, 12*(1), 3-16.

Hanley, N., Wright, R. E., & Adamowicz, V. (1998b). Using choice experiments to value the environment. *Environmental and Resource Economics 11*(3-4), 413-428.

Hanley, N., Wright, R. E., & Alvarez-Farizo, B. (2006). Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive. *Journal of Environmental Management, 78*, 183-193.

Hanley, N., Wright, R. E., & Koop, G. (2002). Modelling recreation demand using choice experiments: climbing in Scotland. *Environmental and Resource Economics, 22*, 449-466.

Hausman, J., & McFadden, D. (1984). A specification test for the multinomial logit model. *Econometrica, 52*, 1219-1240

Heiner, R. A. (1983). The origin of predictable behavior. *American Economic Review, 73*(4), 560-595.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: a primer*. Cambridge, UK: Cambridge University Press.

Holmes, T. P., & Adamowicz, W. L. (2003). Attribute-based methods. In P. A. Champ, K. J. Boyle & T. C. Brown (Eds.), *A primer of nonmarket valuation* (Vol. 3, pp. 171-219). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Huber, J., & Zwerina, K. (1996). The importance in utility balance in efficient choice designs. *Journal of Marketing Research, 33*, 307-317.

Itaoka, K., Saito, A., Krupnick, A., Adamowicz, V., & Taniguchi, T. (2006). The effect of risk characteristics on the willingness to pay for mortality risk reduction from electric power generation. *Environmental & Resource Economics, 33*(371-398).

Jin, J., Wang, Z., & Ran, S. (2006). Comparison of contingent valuation and choice experiment in solid waste management programs in Macao. *Ecological Economics, 57*, 430-441.

Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*, 132-157.

Louviere, J., & Hensher, D. (2000, 2-7 July). *Combining sources of preference data*. Paper presented at the 9th International Association for Travel Behavior Research Conference (IATBR), Gold Coast, Queensland, Australia.

Louviere, J., & Hensher, D. A. (1982). On the design and analysis of simulated choice or allocation experiments in travel choice modelling. *Transportation Research Record, 890*, 11-17.

Louviere, J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research, 20*, 350-367.

Louviere, J. J. (2001). Choice experiments: an overview of concepts and issues. In J. Bennett & R. Blamey (Eds.), *The choice modelling: an approach to environmental valuation*. Cheltenham, UK: Edward Elgar.

Louviere, J. J., Hernsher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and application.* Cambridge, UK: Cambridge University Press.

Luce, R. D. (1959). *Individual choice behavior: a theoretical analysis*. New York: John Wiley & Sons.

Mackanzie, J. (1993). A comparison of contingent preference models. *American Journal of Agricultural Economics, 75*, 593-603.

Mallawaarachchi, T., Blamey, R. K., Morrison, M. D., Johnson, A. K. L., & Bennett, J. W. (2001). Community values for environmental protection in a cane farming catchment in Northern Australia: A choice modelling study. *Journal of Environmental Management, 62*(3), 301-316.

Marschak, J. (1960). Binary choice constraints on random utility indicators. In K. Arrow (Ed.), *Stanford symposium on mathematical methods in the social sciences*. Stanford, CA: Stanford University Press.

Mazzanti, M. (2003). Discrete choice models and valuation experiments. *Journal of Economics Studies, 30*(6), 584-604.

Mazzotta, M. J., & Opaluch, J. J. (1995). Decision making when choices are complex: a test of Heiner's hypothesis. *Land Economics, 71*(4), 500-515.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.

McGonagle, M. P., & Swallow, S. K. (2005). Open space and public access: A contingent choice application to coastal preservation. *Land Economics, 81*(4), 477-495.

Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods. The contingent valuation method.* Washington D. C.: Resource for the Future.

Morey, E. R., Buchanan, T., & Waldman, D. M. (2002). Estimating the benefits and costs to mountain bikers of changes in trail characteristics, access fees, and site closures: choice experiments and benefits transfer. *Journal of Environmental Management, 64*, 411-422.

Morrison, M. D., Bennett, J., & Blamey, R. (1999). Valuing improved wetland quality using choice modeling. *Water Resources Research, 35*(9), 2805-2814.

Morrison, M. D., Blamey, R., Bennett, J., & Louviere, J. (1996). *A comparison of stated preference techniques for estimating environmental values* (No. 1). Canberra: Department of Economics and Management, ADFA.

Othman, J., Bennett, J., & Blamey, R. (2004). Environmental values and resource management options: a choice modelling experience in Malaysia. *Environment and Development Economics, 9*, 803-824.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy and selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 534-552.

Rae, D. A. (1983). The value to visitors of improving visibility at Mesa Verde and Great Smoky National Parks. In R. D. Rowe & L. G. Chestnut (Eds.), *Managing air quality and scenic resources at national parks and wilderness areas*. Boulder, CO: Westview Press.

Riera, P., & Mogas, J. (2004). Finding the social value of forests through stated preference methods: a Mediterranean forest valuation exercise. *Silva Lusitana*, 17-34.

Rolfe, J., Bennett, J., & Louviere, J. (2000). Choice modelling and its potential application to tropical rainforest preservation. *Ecological Economics, 35*(2), 289-302.

Rolfe, J., Bennett, J., & Louviere, J. (2002). Stated values and reminders of substitute goods: Testing for framing effects with choice modelling. *Australian Journal of Agricultural and Resource Economics, 46*(1), 1-20.

Scarpa, R., Ruto, E. S. K., Kristjanson, P., Radeny, M., Drucker, A. G., & Rege, J. E. O. (2003). Valuing indigenous cattle breeds in Kenya: an empirical comparison of stated and revealed preference value estimates. *Ecological Economics, 45*, 409-426.

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research, 29*(3), 281-295.

Swait, J., & Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research, 30*, 305-314.

Swait, J. D., & Adamowicz, W. (1996). The effect of choice environment and task demand on consumer behavior. Paper to 1996 Canadian Resource and Environmental Economics Study Group, Montreal.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34*, 273-286.

Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Willis, K. G., McMahon, P. L., Garrod, G. D., & Powe, N. A. (2002). Water companies' service performance and environmental trade-offs. *Journal of Environmental Planning and Management, 45*(3), 363-379.

ANNEX

Table I. Summary of the most common features of CE applications

Valuation approach	Valuation technique	Study object	Task characteristics					Estimation models
			Number of attributes[a]	Number of attribute levels[b]	Design criterion	Choice set size	Number of replications	
Recreational site choice (use value approach)	Joint models (TCM+CE)	Recreational opportunities Recreational hunting	From 6 to 11[c]	From 2 to 4 (cost proxied by travel distance)	Orthogonality	3 alternatives	16	CL
	Comparison CVM/CE	Recreational hunting	6	From 2 to 4 (cost proxied by travel distance)	Orthogonality	3 alternatives	16	BL CL
	Comparison TCM/CE	Recreational climbing	6	From 2 to 4 (6 cost levels)	Orthogonality	3 alternatives	4/8	CL (basic vs. extended) NL (basic vs. extended)
	CE	Recreational deer hunting Recreational mountain biking	From 5 to 6	From 2 to 3 (cost levels from 3 to 9)	Orthogonality	2 alternatives	From 5 to 6	BL CL BT
Environmental management programs (non-use value approach)	Comparison CVM/CE[d]	Woodlands Wildlife Lanscape Biodiversity Solid waste Soil conservation	From 4 to 6	From 2 to 4[e] (cost levels from 4 to 8)	Orthogonality Utility balance	From 2 to 3[f]	From 4 to 8	CL (linear vs. quadratic/ basic vs. extended)) ML (linear vs. quadratic) RPL

	Comparison CR/CE	Forests Water supply project	6	From 2 to 4 (4 cost levels)	Orthogonality	3	4	CL Rank ordered logit
	Comparison Hedonic/CE	Cattle traits	6	2 (2 cost levels)	Orthogonality	3	8	CL ML ML panel version
	CE	Wetlands Woodlands Rainforests Water ecological status Water supply options Agrobiodiversity Coastal land	From 4 to 10[g]	From 1 to 7[h] (cost levels from 3 to 5)	Orthogonality D-optimal design[i]	From 2 to 3[j]	From 4 to 9	Basic CL (linear vs. logarithmic/ basic vs. extended) NL RPL (basic vs. extended) Latent class model BT

[a] It includes the cost attribute.

[b] The number of levels of the cost attribute tends to be higher than the one for the other characteristics in order to obtain enough variability to estimate WTP. For this reason, it is specified in parenthesis.

[c] Adamowicz et al. (1994)'s work is one of the few CE applications using a high number of attributes (from 10 to 11).

[d] Adamowicz et al. (1998a)'s paper also shows a pooled model estimation. However, the main objective of the paper is to make a comparison between the CVM and the CE.

[e] Adamowicz et al. (1998a) are the only ones using 4 levels for each attribute. The levels for the remaining applications comparing the CVM and the CE range from 2 to 3.

[f] Jin et al. (2006) are the only ones that use 2 alternatives in the choice sets.

[g] The study that uses 10 attributes is the one carried out by McGonagle & Swallow (2005). However, it is to recall that from the 10 attributes, the site attributes are only 6.

[h] The attribute with 7 levels is used in the study done by Rolfe et al. (2000). It is not an ecological attribute but an attribute indicating *location*. The number of levels for the rest of attributes in CE applications ranges from 1 to 4.

[i] The D-optimal design is only used in Carlsson et al. (2003).

[j] The only study in which there are only 2 alternatives in the choice sets is the one carried out by Rolfe et al. (2000).